Ecological and epidemiological consequences of rapid urbanisation at wildlife-livestock-human interfaces

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by

James Mark Hassell

March, 2018

Abstract

Ecological and epidemiological consequences of rapid urbanisation at wildlife-livestock-human interfaces

Urbanization is characterized by rapid intensification of agriculture, socioeconomic change, and ecological fragmentation, which can have profound impacts on the distributional ecology of host populations and epidemiology of infectious disease within them. In this thesis, results from a large-scale field study conducted in Nairobi, Kenya are used to explore how anthropogenic and ecological changes associated with urbanisation influence the structure of sympatric wildlife, livestock and human host populations, and dictate bacterial epidemiology in wildlife hosts. As likely points of contact (and thus parasite transmission) between vertebrate wildlife, livestock, and humans, household 'interfaces' were chosen as sampling units. The ecological and sociological status of households was characterised through ecological surveys, questionnaire data and geospatial mapping, and faecal samples were collected from wildlife occurring within the household compound, and livestock and human inhabitants. *Escherichia coli* was isolated from faecal samples, and characterised both phenotypically (through antimicrobial sensitivity testing) and genetically (through whole genome sequencing).

In the first part of this thesis I consider the influence of urban land-use change on the structure of host populations at household interfaces. Using unsupervised machine learning I describe variation in the host composition of wildlife-livestock-human interfaces and, through multivariate regression analysis, demonstrate that city-wide variation in ecological and anthropogenic factors (such as biotic habitat diversity and wealth) drive structural changes in wildlife, livestock and human populations across the urban landscape of Nairobi.

Utilising commensal *E. coli* as an exemplar organism, I proceed to explore epidemiological connectivity between wildlife, livestock and humans at household interfaces, and link epidemiological processes in urban wildlife to their drivers across the urban landscape. Firstly, I explore the epidemiology of clinically relevant antimicrobial resistant (AMR)-*E. coli* in urban wildlife in Nairobi. Comparing *E. coli* isolates in wildlife to livestock, humans and the environment, I find that *E. coli* isolated from wildlife have a lower diversity of resistance phenotypes, and are thus an unlikely source of AMR. At household interfaces, I find evidence of AMR-*E. coli* exchange between rodents/seed-eating birds, and cattle and humans, and demonstrate that transmission is facilitated through anthropogenic resource provision in households. Next, utilising high resolution sequencing data, I explore the response of microbial communities in wildlife hosts to urban land-use change. Specifically, I test the hypothesis that communities of bacterial mobile genetic elements (MGEs) are deterministically structured, according to changes in host community structure. I show that the diversity of genes encoding virulence and AMR in avian-borne *E. coli* is determined by variation in the distribution and density of birds, livestock and humans at household interfaces, and that this varies along gradients of urbanisation. To conclude, I relate the findings in this thesis across multiple scales, linking the influence of abiotic factors such as habitat alteration and socioeconomics to host community structure at household interfaces and the epidemiology of wildlife-borne *E. coli*. Using this framework, I suggest future directions for research on urban disease emergence, and discuss implications of my findings for public health and urban planning.

Acknowledgements

This PhD was funded by the UK Medical Research Council, through the Environmental and Social Ecology of Human Infectious Diseases Initiative (ESEI), a cross research council initiative supported by the Medical Research Council, Biotechnology and Biological Science Research Council, the Economic and Social Research Council and the Natural Environment Research Council (grant reference: G1100783/1). I also received support for field activities from the CGIAR Research Program on Agriculture for Nutrition and Health (A4NH), led by the International Food Policy Research Institute.

I should like to begin by thanking my supervisors, Eric Fèvre, Mike Begon and Melissa Ward. I feel particularly indebted to Eric, who provided me with the opportunity to undertake a PhD in Kenya, and has offered me the guidance, resources and freedom to really make this project my own. I am hugely thankful to Melissa for sharing her expertise, and for the many meetings and discussions that have proven invaluable to my research. I shared many productive conversations with Mike, which were integral in pursuing a focus on ecological change at epidemiological interfaces.

I thank everyone in Eric's zoonotic and emerging diseases group (past and present) for their friendship, assistance and support over the last four years, and for helping make Kenya my home. Judy Bettridge, in particular, has been a source of epidemiological and statistical wisdom, and I have spent many hours bouncing ideas off Dishon Muloi. I thank my part-time office mates Laure Madé, Maud Carron, Annie Cook and Stella Kiambi, for their stimulating discussion and good humour throughout. My family are responsible for nurturing my fascination with wildlife, Africa and science, and for this, alongside their constant support and encouragement, I will be forever grateful.

Finally, I would like to thank Victoria, who has put up with so much, and been there for me throughout.

Declaration

I declare that this thesis and the analyses described within it are original and my own composition, expect where explicitly stated below. Data collection described herein was primarily carried out by myself and other members of the UrbanZoo team, following protocols authored in whole or in part by myself (further details are given in Chapter 3.1). It has not been submitted for any degree or professional qualification except as specified.

James M. Hassell

15 March, 2018

Chapter 5

Bacterial culture and sensitivity testing were conducted by laboratory teams at the University of Nairobi (UoN) and Kenya Medical Research Institute (KEMRI) (see Chapter 3.1).

R code for performing comparative diversity analysis of antimicrobial resistance profiles between wildlife, livestock, humans and the environment was kindly provided by Dr. Bram Van Bunnick, who adapted it from the original code used by Mather *et al.* [1].

Chapter 6

Bacterial culture and *Escherichia Coli* DNA extraction were conducted by laboratory teams at UoN, KEMRI and the International Livestock Research Institute (see Chapter 3.1).

Whole Genome Sequencing (WGS) of *E. coli* isolates was performed by the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics in Oxford (funded by Wellcome Trust grant reference 090532/Z/09/Z). Bioinformatic processing of sequences was conducted by members of the Modernizing Medical Microbiology Informatics Group, Dr. Melissa Ward and Dishon Muloi.

Contents

List of Figures

List of Supplementary Figures

List of Tables

List of Supplementary Tables

Chapter 1

Introduction

1 Introduction

1.1 Motivation

It is widely accepted that urbanisation, particularly when associated with social disparity (and resulting health inequalities), large scale migration, poor living conditions, and close contact with domestic animals, poses a risk to human health through the emergence and spread of infectious disease [2]. Cities in low-middle income countries, which are characterised by rapid, unplanned urbanisation [2], are thought to be particularly at risk. This is especially true of urban and peri-urban zones in Africa, where growth and migration is expected to result in an increase in the population residing in these areas from 35% in 2007, to 51% by 2030 [3]. Without adequate planning, fast rates of urban growth have knock-on effects on human policy-orientated fields such as health provisions and delivery, sanitation, demographics, trade, economics and food production, whilst resulting in substantive environmental change. Considerable attention has been paid to the impacts of urbanisation on these disciplines, but there has been minimal effort to link environmental changes to human social organisation, particularly when assessing their combined effects on the transmission of pathogens. Whilst rapid urbanisation probably increases the likelihood of pathogen emergence, the underlying mechanisms are poorly understood, studied and quantified.

Approximately 60% of human pathogens are zoonotic [4], whilst 80% of novel pathogens have zoonotic origins [5], and as such, both wild and domestic animals are a key source of emerging diseases in humans. In Africa, urban environments act as a crucible for interactions between humans, animals, and their shared environment. Informal livestock keeping is a common feature of major African cities (22% of all urban households keep livestock in Kampala, Uganda [6], whilst 50% of households in parts of Nairobi, Kenya, keep one or more forms of livestock [7]), where poor management of livestock and human waste products can contaminate the environment, and provide resources that attract urban wildlife. Fragmentation of the biologically rich natural habitats in which these wildlife exist, can cause dramatic changes to the composition of wildlife populations, with implications for the biology of their pathogens, and interactions with humans and livestock [8]. This degree of mixing and

contact between wildlife, livestock and humans creates diverse ecological niches (or 'interfaces'), which present broad opportunities for pathogen transmission, and disease emergence across urban landscapes [9]. Studies linking urban ecological systems and human social organisation to disease emergence are severely lacking, and for these interfaces to be of relevance to those responsible for mitigating the occurrence of disease emergence in urban settings, the risks that they pose to human health must be understood. As such, the range of forms that wildlife-livestock-human interfaces take, the environmental or anthropogenic determinants for their ecological structure, and their epidemiological importance, need to be fully characterized.

1.2 Aims

In this thesis, I set out to develop, and test, key hypotheses aimed at understanding how rapid urban change in Nairobi, Kenya (consisting of variation in urban land use, and human social organisation), leads to the formation of wildlife-livestock-human interfaces, and influences the epidemiology of pathogen transmission between hosts. In this respect, my thesis can be split into three components; *i)* a review of existing evidence for the drivers and epidemiology of emerging wildlife-borne zoonoses in urban landscapes, in which hypotheses are generated for the study of pathogen emergence at complex multi-species urban interfaces; *ii)* ecological hypothesis testing, which is used to understand the form that household wildlife-livestock-human interfaces take across the city of Nairobi, and determinants for variation in the structure of multi-species interfaces; *iii)* epidemiological hypothesis testing, utilizing antimicrobial resistance (AMR) phenotypes and genetic derivatives of *Escherichia coli* as an exemplar of many potential emerging pathogens, which is used to investigate routes of bacterial transmission between sympatric wildlife, livestock and humans at household interfaces, and the influence of urbanisation on bacterial epidemiology in wildlife communities.

1.3 Chapter outline and hypotheses

In Chapter 2, I conduct a thorough review of existing evidence for the drivers and epidemiology of emerging wildlife-borne zoonoses in urban landscapes, generating a major synthesis on the ecological drivers of disease emergence in rapidly urbanising

settings, and new insight into the importance of urban wildlife-livestock-human interfaces.

In Chapter 3, I describe my role in the design and execution of the UrbanZoo 99-household study, a major field study conducted between September 2015 and September 2016, aimed at investigating the epidemiology and ecology of disease emergence in Nairobi. The samples and metadata collected from households in this study are used to test ecological and epidemiological hypotheses in Chapters 4-6.

In Chapter 4, I use observational data collected as part of the UrbanZoo 99-household study to test the following ecological hypotheses, aimed at understanding variation in the form of household wildlife-livestock-human interfaces across Nairobi:

- Communities of sympatric wildlife, livestock and humans form predictable assemblages, with comparable ecological structure, across the urban landscape.

- Urban drivers (such as variation in land use and human sociological factors) are associated with the structure of wildlife and livestock host communities.

In Chapters 5 and 6, I use *E. coli*, collected from sympatric wildlife, livestock and humans in Nairobi, as a model organism to test epidemiological hypotheses aimed at understanding how urbanisation influences bacterial transmission at household wildlife-livestock-human interfaces, and the epidemiology of bacterial diseases in wildlife populations across the city.

In Chapter 5, *E. coli* AMR phenotypes are used to test the following hypotheses aimed at investigating urban determinants for bacterial exchange between sympatric wildlife, livestock and humans:

- Wildlife are a net recipient (or 'sink') of *E. coli* AMR phenotypes in urban environments, when compared to sympatric livestock, humans and their shared environment.

- Exposure of wildlife to *E. coli* AMR phenotypes in urban environments is determined by taxa-specific functional ecology (e.g. feeding traits).

- Exchange of *E. coli* AMR phenotypes occurs between sympatric wildlife and livestock, but is dependent upon livestock-keeping practices, and the functional ecology of wildlife taxa.

In Chapter 6, high resolution *E. coli* genetic data (derived from whole genome sequencing) is used to test the following hypotheses, aimed at describing the response of microbial genetic communities in the wildlife population to urbanisation, and the population genetic structure of *E. coli* at household interfaces.

- Determinism in the structure of communities of virulence and AMR genes in wildlife-borne *E. coli*, is associated with changes in wildlife assemblages, and anthropogenic factors (such as livestock-keeping practices) respectively.
- Genetic differentiation between *E. coli* collected from wildlife, livestock and human hosts at household interfaces is driven by geographic distance between hosts, rather than host taxonomic identity.

In Chapter 7, I summarise the main findings of the thesis, bringing together data to form a picture of how urbanisation influences the ecology and epidemiology of wildlife-livestock-human interfaces. I also make recommendations for future research areas and the potential for policy interventions to improve human and animal health in rapidly developing urban environments.

Chapter 2

Review of urbanisation and disease emergence at the wildlife-livestock-human interface

The published manuscript resulting from this chapter can be accessed from the following citation: *Hassell, J. M., Begon, M., Ward, M. J., & Fèvre, E. M. (2017). Urbanization and disease emergence: Dynamics at the wildlife–livestock–human interface. Trends in Ecology & Evolution, 32, 55–67.*

2 Introduction

Urbanisation in developing countries is characterised by rapid intensification of agriculture, socioeconomic change, and ecological fragmentation, which can have profound impacts on the epidemiology of infectious disease. In this chapter, current scientific evidence for the drivers and epidemiology of emerging wildlife-borne zoonoses in urban landscapes, where anthropogenic pressures can create diverse wildlife-livestock-human interfaces, are reviewed. These interfaces represent a critical point for cross-species transmission (and emergence of pathogens into new host populations), and thus understanding their form and function is necessary to identify suitable interventions to mitigate the risk of disease emergence. To achieve this, interfaces must be studied as complex, multi-host communities whose structure and form are dictated by both ecological and anthropological factors.

2.1 Emerging diseases in changing landscapes

Emerging infectious diseases (EIDs) (see Glossary) are recognised as pathogens 'whose incidence in humans has increased within the past two decades or threatens to increase in the near future' [10]. As well as describing the spread of newly evolved or previously undetected pathogens, and pathogens that are increasing their geographic spread, increasing their impact, changing their clinical presentation or moving into human hosts for the first time, the term 'emergence' may also be used to describe the reappearance (or 're-emergence') of a known infection after a decline in incidence [10]. It is estimated that between 60 – 80% of newly emerging infections are zoonotic in origin and thus are (at least initially) dependent on an animal reservoir for survival [5,11]. Of these emerging zoonoses, at least 70% have a wildlife origin, with cross-species spread and onward transmission representing a natural response to the evolutionary pressures of pathogen ecology [11,12]. Whilst both wildlife and domesticated animal reservoirs can be considered important sources of EIDs, it is the anthropogenic influence on ecological systems that dictates the level of risk that operates at the interface between humans and animals in zoonotic disease emergence.

Interactions between humans and the ecosystems in which they exist have occurred for as long as there have been humans. However, over the past 10,000 years, human-ecosystem interactions have become increasingly profound following a series of chronological transitions: *i*) the establishment of *local* settlements, agriculture and domestication of livestock, *ii*) *regional* contact through trade, *iii*) *intercontinental* exploration, imperialism and industrialisation, and *iv*) *globalisation*, urbanisation and climate change [13]. Current levels of human-ecosystem interaction, driven by increased environmental encroachment and land-use change (exploitation of natural resources and agricultural practices), and environmental effects such as climate change, may result in habitat alteration and changes in species assemblage and contact rates that promote the emergence of zoonotic disease. It is estimated that spread and persistence of newly emerged (or re-emerged) pathogens may then be perpetuated by a combination of factors including expanding global human populations and urbanisation, international trade and travel, intensive livestock keeping systems, proliferation of reservoir populations, and antimicrobial drug use [12,14–16]. Land-use change, through population-driven anthropogenic influences such as forestry, mining, agriculture, and urban and industrial development, is frequently associated with disease emergence [17,18].

Urbanisation can be considered a key driver of land-use change that is likely to increase at an unprecedented rate in the coming decades, particularly in developing countries, where as much as 90% of population growth is projected to occur in cities [19,20]. Human population density and growth are significant predictors of historical EID events, and thus urbanisation is likely to have a profound effect on public health as rural pathogens adapt to urban conditions, and other pathogens emerge (or re-emerge) in urban areas [11]. Human factors such as population density, migration, trade, sanitation and access to clean water may promote the transmission of pathogens and alter vector dynamics, whilst social factors that drive health inequality (socioeconomic status, housing, race, ethnicity, gender and education) also influence the epidemiology of infectious disease in urban areas [2,8] (Figure 1.1). In this thesis, I focus on rapid urbanisation (predominantly a feature of developing countries) as a driver of disease emergence, and use it to explore how anthropogenic changes are driving interactions and the potential for disease emergence between sympatric wildlife, livestock and humans.

2.2 Urbanisation and disease emergence

Spatial overlap between hosts is a key requirement for disease emergence. As such, in order to investigate the conditions in which urbanisation may lead to the emergence of zoonotic disease across species, and thus risk factors for transmission to humans, it is necessary to simplify the complexity of urban systems by considering them as a network of interfaces across which pathogens can be transmitted; the physical interfaces at which humans and animals interact and pathogens are exchanged exist within the context of societal and policy interfaces (as depicted in the schematic in Figure 1.1). These networks exist at different scales. At a local-scale, households form part of what can be considered urban communities (or 'metapopulations'); groups of similar physical interfaces that are characterised by a set of societal (e.g. demographic and socioeconomic) characteristics. These metapopulations are linked by networks of connectivity that span the urban landscape, such as movement of people, livestock and their products, and wildlife, and the environment [21,22]. As a result, key drivers that could promote interaction between humans and animals are: *i*) livestock keeping practices, production systems and the movements of livestock and animal products in urban areas, and *ii*) the direct effects of urbanisation on the physical environment, ecosystems in which urban centres are developed, and animal communities that exist freely within these. Urban systems are highly complex and the factors listed above are likely to influence the type and extent of human interactions with livestock, animal products and ecosystems, resulting in the creation of human-animal 'interfaces' which may promote the transmission of disease between animals and people.

Urban-adapted (referred to here as 'synanthropic') wildlife is abundant in cities, and is composed of species that can respond to behavioural and resource-based selection pressures imposed by urban environments [23]. Many synanthropic species have been shown to carry zoonotic pathogens and in some cases act as reservoir hosts for these pathogens. Studies generally focus on those species that are found ubiquitously within human environments and that commonly act as hosts for zoonotic diseases, such as rodents, birds, bats and certain other species of mammal (e.g. foxes in Europe and racoons in the US) [24,25]. Rodents, for example, harbour important zoonoses such as plague, leptospirosis and hantaviruses, and the emergence and re-

emergence of these pathogens in human populations is seemingly linked to increasing urbanisation and urban poverty in developing countries and the ecology of zoonotic pathogens in rat populations [26–29]. Anthropogenic changes associated with urbanisation can also bring bats into closer contact with livestock and humans and alter disease ecology [30,31]. As such, human activities that increase exposure to populations of urban-dwelling wildlife species will undoubtedly increase the risk of pathogens spilling over into humans or livestock, but little is known of the epidemiological processes by which this occurs at such interfaces.

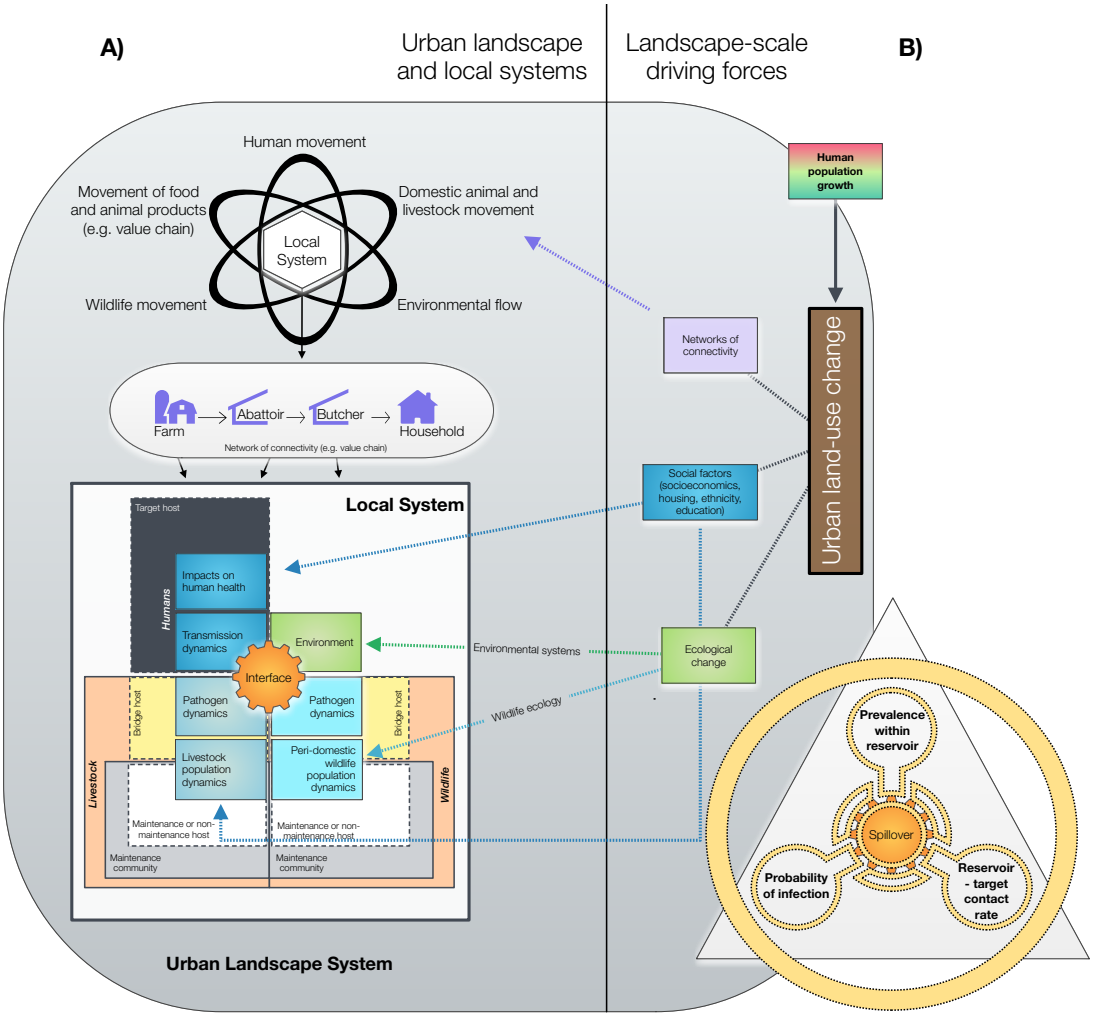


Figure 2.1. A conceptual framework for disease emergence in urban landscapes (adapted from Viana *et al.* [61] and Lloyd-Smith *et al.* [40]). **A)** This framework incorporates urban land-use change and its effects on two spatial scales: at a systems and local level. A simplified disease reservoir framework is included at the local level, in which livestock and synanthropic wildlife may exist within the maintenance community as maintenance hosts (populations within the reservoir that can maintain the pathogen) or non-maintenance hosts (populations within the reservoir that cannot maintain the pathogen, therefore acting as 'vectors'), or as bridge hosts that exist outside the maintenance community. **B)** Following Lloyd-Smith et al. [44], spillover, which in this framework may relate to pathogen transfer in all directions except for target to reservoir, is governed by the force of infection consisting of the three elements shown.

2.3 Epidemiology and the wildlife-livestock-human interface

Most infectious diseases circulate in communities composed of hosts that are infected with multiple parasites, and parasites that can infect a variable diversity of hosts. Small changes in parasite community structure (within-host competition, or perturbations from host population dynamics) can result in far-reaching consequences for epidemiology of multi-host and single host (monoxenous) parasite species [32–34]. Such downstream epidemiological effects are demonstrated in a number of well-studied multi-host zoonotic disease systems, including the seasonal and co-infection dynamics of Cowpox virus [35], Lyme disease in white-footed mice [36], and Nipah and Hendra virus in fruit bats [30,37,38]. With the emergence of high profile pathogens that exhibit wide host plasticity (such as Ebola virus and Avian influenza), a community approach is being increasingly embraced for studying the multi-host ecology of zoonotic pathogens.

Studying the role of wildlife in multi-host disease systems is complicated by ecological and behavioural attributes unique to these species, and by the influence of natural and human systems, both of which complicate conceptual models of disease transmission [39]. Following the disease reservoirs framework recently revised by [40] and [41], in a multi-host pathogen system where wildlife either exist within the maintenance community as a maintenance or non-maintenance host, or outside the maintenance community as a bridge host, the dynamics of a zoonotic disease involve two phases: *i)* transmission between maintenance and/or non-maintenance host species (wildlife and/or domestic) within the reservoir, and *ii)* spillover transmission to humans from the maintenance community (Figure 1). In basic models, the persistence required for hosts to maintain a zoonotic pathogen and thus act as a maintenance community can be described by the basic reproductive number (R_0: the transmission potential of a pathogen), whilst risk of spillover transmission to humans is defined by the force of infection from animals to humans. Contact is a key feature of both reservoir and disease emergence dynamics. Thus, R_0 is closely linked to the rate of contact between susceptible and infectious individuals and the recovery rate of infected individuals, and the force of infection (and thus risk of human spillover) is determined by prevalence of infection in the maintenance population and/or bridge host(s), the rate of contact between humans and infected individuals, and the

probability that infection occurs upon contact [42–44]. However, host ecological traits (such as life-history characteristics, seasonality, coloniality and sympatry) and population-level changes brought on by land-use change are likely to play a large role in pathogen transmission and persistence in wildlife and livestock species [39,45]. These factors (particularly human ecology) will strongly influence contact between wildlife, livestock and humans, and prevalence of infection in animal reservoirs, and are therefore of fundamental importance to reservoir dynamics and disease emergence in changing landscapes.

Murray and Daszak [46] discuss two conceptual models for disease emergence under land-use change; the 'perturbation' and 'pathogen pool' hypotheses. The pathogen pool hypothesis assumes exposure to novel diseases from a diverse pool of pathogens in wildlife to which humans or livestock, as naïve hosts, have not had prior exposure. The perturbation hypothesis focuses on a more dynamic model for disease emergence, where land-use change forces perturbations in disease dynamics within the reservoir, before emergence occurs in humans or livestock. In reality, it seems unlikely that these two hypotheses are mutually exclusive; evidence from empirical studies generally favours a dynamic model for disease emergence [47]. As such, the extent to which 'perturbation' (changes in species richness, abundance and contact rate) or the 'zoonotic pathogen pool' dictate risk of emergence at urban interfaces, is probably dependent on the impact of urbanisation on community ecology, and the degree of co-evolution between sympatric wildlife, humans and livestock at each interface.

2.4 Influence of urbanisation on pathogen dynamics within multi-host wildlife systems

Associations between urbanisation and the prevalence of diseases in populations of free-ranging wildlife have been described for a wide taxonomic range of host species and pathogens (reviewed by Bradley and Altizer [8]. Evidence suggests that through altered habitat structure and changes to resource availability, urbanisation results in significant changes to the structure of wildlife communities, which are subsequently characterised by low biodiversity with proportional increases in abundance of certain 'generalist' species [48,49]. From a landscape-scale perspective, this results in a declining trend in species richness from rural areas to urban centres

("biotic homogenisation") with synanthropic species occurring at higher densities in urban and suburban environments than less-disturbed areas [8,50,51]. Not surprisingly, such profound changes in trophic structure will have epidemiological consequences for pathogens within these communities, and as a general rule, declining host biodiversity should be matched by a loss in parasite diversity, thus reducing the 'pathogen pool' and with it the risk of novel disease emergence [52]. However, the epidemiological consequences of changes to such a system are likely to be pathogen-specific, and dependent on how trophic reassortment affects the following parameters: likelihood of encounter and transmission between competent hosts, host abundance/density, and infected host mortality and recovery [53]. For example, helminth species richness of rodents in South East Asia is positively associated with decreasing rodent species richness, and increasing rodent abundance and level of synanthropy [54]. Increases in synanthropic species population density may elevate contact rates (through changes in host ranging patterns and densities), and thus increase the risk of pathogen transmission via direct contact and oro-faecal routes. On the other hand, fragmentation of these populations may result in genetic bottlenecks and subsequently reduced effective immune responses [43,51,55,56]. As host diversity decreases along gradients of urbanisation, many pathogens are lost, but some (notably those in the hosts that remain in low diversity communities) may increase as a result of increased host abundance [36,57]. Reverse zoonotic transmission (zooanthroponosis) from humans to wildlife may also pose a threat to wildlife populations with increased exposure to humans [58,59]. The epidemiological effects of urbanisation can therefore have important implications for both wildlife conservation and public health, with marginal wildlife species being susceptible to infection with pathogens circulating in urban-adapted hosts, and the potential for increased circulation of certain zoonotic disease in competent synanthropic reservoir hosts.

2.5 Interfaces between sympatric wildlife, livestock and humans in an urban landscape

Heterogeneous patterns of wildlife distribution are a feature of urban landscapes, and certain species group in spatial aggregations with livestock (or their products) and humans, creating interfaces that may be important for the transmission

of zoonotic disease. As described, the dynamics of infection at these interfaces are determined by changes in diversity, abundance and contact rates between reservoir and target hosts, thus influencing risk of cross-species disease transmission. Several systematic reviews have identified 'high-risk' interfaces for zoonotic disease transmission on a global scale; specific interfaces for spillover from wildlife include human dwellings, agricultural fields and occupational exposure, whilst broader descriptions include agricultural intensification and environmental change [16,60]. However, as argued by Jones at al. [16], attempts to describe systems within which pathogens emerge or change in virulence have predominantly focused on global generalisations, which may not be appropriate to capture the heterogeneity of interfaces. Instead, interfaces and the driving factors that define them should be studied at appropriate, spatially explicit scales [61]. These feedback loops are considered at a hypothetical urban wildlife-livestock-human interfaces in Box 1.

From an ecosystem perspective, anthropogenic pressures result in the fragmentation of natural biomes, leaving a composite mix of different habitats. Remnant fragments that are representative of the original biome may be thought of as 'patches' that exist within a 'matrix' of habitats that are unlike the original [75–77]. Interfaces between patches and the matrix exist at local scales, and may be classified as 'ecotones' – edges or transitionary zones between adjacent ecological systems where "biophysical factors, biological activity and ecological evolutionary processes are concentrated and intensified" [78]. It has been suggested that by expanding ecotonal areas through interspersing human landscapes such as farmland and settlements with natural landscapes, anthropogenic influences may alter disease 'niches' by bringing together humans, vectors and reservoir hosts (wildlife or domestic animals), thus increasing contact and the risk of transmission [76]. Such landscape changes can be compounded by alterations in wildlife species interaction and abundance (e.g. host ecological traits). For example, rodents may undergo 'ecological release' at forest interfaces being attracted to farmland and human settlements for resources and suitable breeding habitat, and human settlements may provide suitable breeding habitat for mosquitos and birds (important arthropod vectors and reservoirs for West Nile virus) [78,79]. Evidence for an association between disease emergence and ecotones has been documented for several zoonoses with wildlife reservoirs, including yellow fever, Nipah virus encephalitis, influenza, rabies, hantavirus

Box 1. Dynamics of cross-species transmission at an urban interface

In considering how urbanisation could drive the emergence of a directly transmitted zoonotic disease through changes to the distributional ecology of wildlife, it is helpful to consider urban land-use change as a combination of abiotic factors that can influence the biotic niche within which a focal host species and pathogen exist. In a classical ecological sense, abiotic niches are used to describe a set of conditions that permit persistence of a host within a certain geographical range (such as climate), whilst the host species biotic niche is characterised by the inhibitory or facilitatory impact of other species on its existence [62]. Biotic factors are likely to be scale-dependent; ecological and epidemiological processes may operate differently at different scales. For example, the prevalence of rodent associated zoonoses varies widely between and within cities, likely being determined by site-specific abiotic factors such as physical micro-environments [29]. Efforts to understand scale-dependent processes, such as the role of geographical distribution and ecological setting in creating opportunities for pathogen transmission to occur, have led to the emergence of landscape epidemiology – a discipline that incorporates the framework of traditional epidemiology with the field of landscape ecology to facilitate the study of disease in relation to the ecosystem in which it is found and the human population at risk [63]. This field would have application when addressing the complex and profound effects of urbanisation on wildlife population dynamics, and thus the ecology of zoonotic disease in these systems.

At a hypothetical wildlife-livestock-human interface one can imagine a cascade of abiotic and biotic changes creating conditions suitable for disease emergence at different scales. Abiotic factors (which include climate, resource provisioning, pollution and habitat alteration) exist at multiple scales and facilitate or inhibit the survival of new and existing wildlife species within the reservoir community, which dictates the structural assemblage and 'fitness' of hosts (Figure 2). At a finer scale, these factors influence the biotic niche and dynamics of pathogens within the system; abiotic changes can directly influence microbiota (e.g. driving antimicrobial resistance) [64,65], whilst host diversity, density, phylogenetic structure (ecological, physiological and genetic similarity), and immunocompetence and immunological history of individuals all play an important role in host-pathogen interactions [66,67]. For example, not all con-specifics will be competent hosts for a given parasite, and as 'dead-end' hosts may play a role in regulating infection, whilst direct ecological interactions such as predation or competition will affect the population dynamics and distribution of competent reservoir hosts. Poulin [68] and Reperant [69] considered these factors as applied to the theory of island biogeography, where abiotic drivers influence the degree of interactions within source areas (sources of parasites such as wildlife reservoirs) and island areas (the recipient or 'target' hosts), and the source-island distance (interactions between sources of parasites and recipient host populations that may drive spillover).

How these factors pertain to spillover to a target host (such as humans or livestock) is dependent on the force of infection, components of which include characteristics of the target host and the pathogen, and how the pathogen responds to changes in its biotic niche. Because abiotic factors influencing this system are driven by human activity, increased spatial overlap of humans, their livestock and wildlife is likely, but may not be enough to secure cross-species transmission. If the target host is human, then the likelihood of spillover may be moderated by individual and community variables such as social structure, living conditions, economic status, health and 'risky' behaviour. From a pathogen genetic perspective, a jump into a new host species can either be non-adaptive (a process known as 'ecological fitting', where pre-existing traits allow emergence), or may require adaptive change (mutation of the pathogen in the reservoir host or the new host) [70]. Whether infecting new hosts is an adaptive or non-adaptive process depends on characteristics of the pathogen and host-reservoir. Certain pathogens (RNA viruses in particular) seem inherently more adept at taking advantage of new epidemiological opportunities than others, possibly due to high mutation rates and broader host plasticity [60,71]. For others, structural properties of the reservoir may play a greater role and facilitate ecological fitting; phylogenetic distance between target (human) and reservoir hosts is a predictor of successful host jumps, whilst species belonging to certain phylogenetic clades may possess cellular components that make them more susceptible to pathogen invasion, regardless of phylogenetic distance from existing host species [60,72]. [73] showed that a higher density of susceptible hosts leads to an increased mutation rate and thus increased likelihood of viral host jumps occurring, whilst higher levels of inter-species transmission may lead to the adoption of more generalist pathogen virulence strategies [74]. Although conducted in unnatural microcosms, these studies demonstrate that the characteristics of the pathogen's biotic niche, such as competition from sympatric microorganisms and host-reservoir dynamics, may also drive adaptive evolutionary processes. Finally, changes to population size and immune status of the reservoir may drive an increase in pathogen prevalence, thus amplifying the likelihood of spill-over occurring. This leaves a key challenge to classify the levels of dynamic changes in organisational structure at different interfaces.

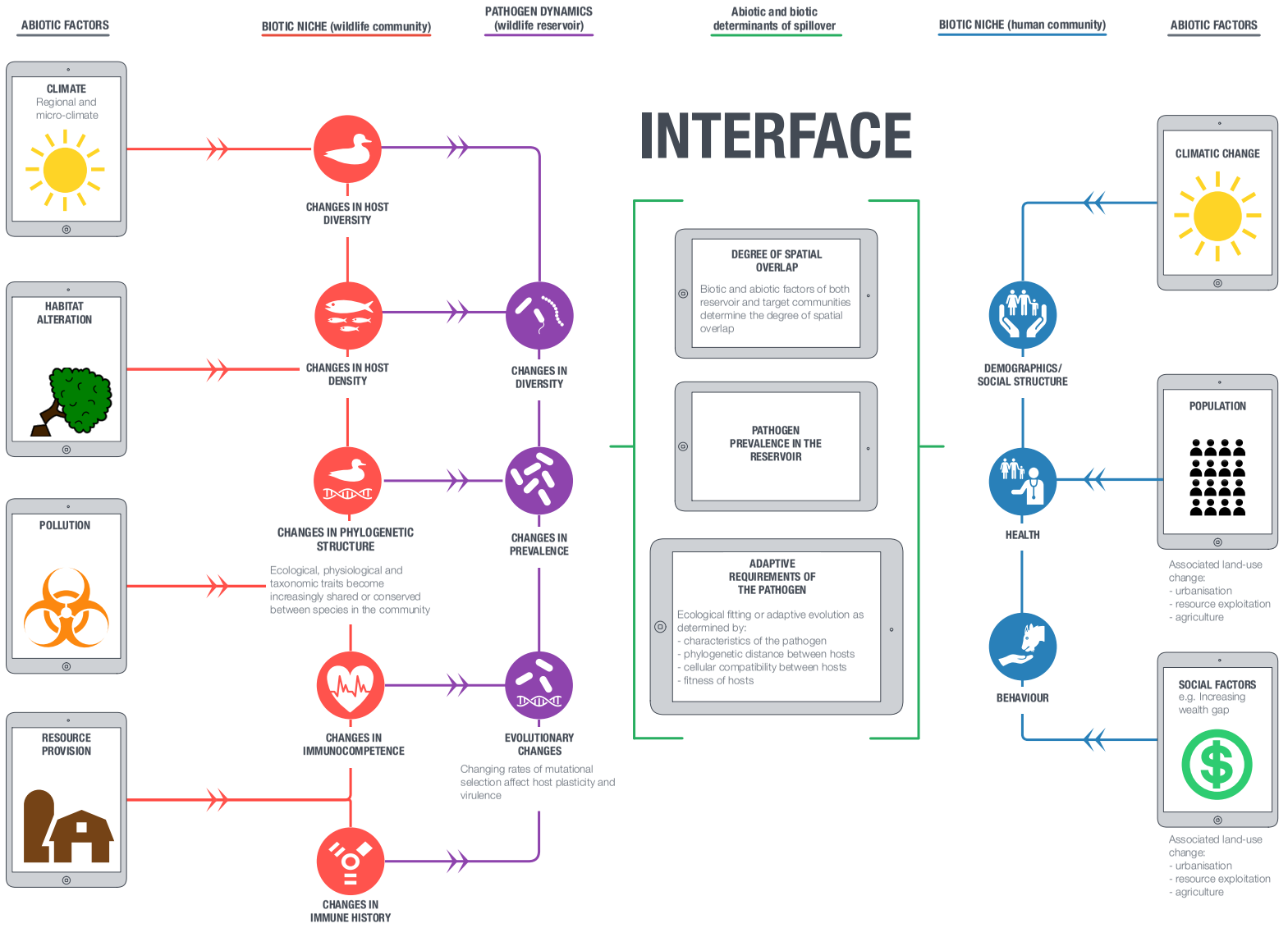


Figure 2.2
Cascades of abiotic factors, and the components of host-pathogen biotic system (for a directly transmitted pathogen) that are affected by these factors, are represented on either side of a hypothetical wildlife-human interface. Biotic systems represented here include a multi-host wildlife community (acting as a parasite reservoir and composed of maintenance and non-maintenance hosts, and non-susceptible species with direct ecological interactions with the reservoir), a human community (acting as variably susceptible 'target' hosts) and the community of parasites within the wildlife reservoir. The requirements for spill-over are represented centrally at the interface.

pulmonary syndrome, Lyme disease, cholera, *E. coli* and African trypanosomiasis [78,80–82]. In urbanised areas such as cities, tangential variation in land use from rural – peri-urban – urban areas would be expected to generate a wide variety of ecotones on micro- and macro- spatial scales. Ecotones may therefore represent important local-scale ecological interfaces within which zoonotic diseases circulate and infect wildlife, domestic animals and humans.

Another important factor in influencing inter-specific wildlife contact and human-livestock-wildlife contact in urban environments is resource provisioning [24,25]. Clumping of resources occurs widely across urban environments at local (e.g. household) and landscape scales, whether as a result of variation in sanitation, refuse and agricultural by-products, livestock keeping practices, supplemental feeding of garden birds or household food availability [8,83–86]. Informal livestock keeping is commonplace in African cities, and often characterised by low biosecurity and mixed-species livestock being kept in close proximity to humans. Evidence from recent zoonotic emergence events [such as Nipah and HPAI (highly pathogenic avian influenza) viruses] and the circulation of relatively stable zoonoses (such as hepatitis B virus and bovine tuberculosis) implicate a role for livestock acting as bridge hosts, epidemiologically linking wildlife and humans [37,87,88]. Whilst resource provisioning commonly leads to increased contact rates between synanthropic wildlife, humans and livestock, pathogen dynamics are also driven by susceptibility to infection, which, depending on the nature of provisioning, may be increased or decreased by host physical condition and immune defence [31,86,89]. In Eastern Australia, the decline in natural food resources, and an abundance of flowering resources in urban gardens has resulted in increasingly large urban colonies of *Pteropus* spp. bats (flying foxes) existing sympatrically with human and horse populations. These bats act as a reservoir for Hendra virus, and have historically lived in widely dispersed, inter-connected metapopulations. Plowright *et al.* [31] demonstrated that the effects of urban development on these metapopulations, through increased contact with humans and horses, and reduced connectivity between flying fox colonies, could dramatically influence the epidemic dynamics of the virus in flying foxes, and increase the risk of Hendra virus emergence in horses and people. Using mechanistic models, Becker et al. [90] and Becker and Hall [86] also demonstrated host demographic, contact and immunological effects of provisioning on R_0, finding that unless provisioning reduces dietary exposure to pathogens or strongly improves

host condition and immunity, increased aggregations of wildlife species dramatically increase pathogen invasion success and long-term prevalence. Environmental stressors such as heavy metal and pesticide pollutants, characteristic of certain urban environments, may further compound these outcomes through their effects on immunological function [91]. As such, resource provisioning is likely to increase host density (a key driver of transmission rates) and wildlife-livestock-human contact, making such areas important interfaces for disease emergence.

Table 1.1 applies the conceptual framework of wildlife-livestock-human interfaces developed by Jones et al. [16] to an urban setting such as Nairobi. Nairobi is a good example of a developing country city with human-livestock-synanthropic wildlife interfaces. It is a city, like many others, that has a growing boundary/edge which makes such contact more likely both on its edges and internally. In this context, urban interfaces created through habitat fragmentation and resource provision are considered. Such clear definition of interfaces is required to simplify the heterogeneous juxtaposition of humans and animals in urban landscapes, and thus enable the application of ecological, epidemiological and anthropological approaches to the study of these landscapes. As well as capturing complex human and ecological processes that underlie disease emergence in urban landscapes, by studying these interfaces along rural – peri-urban – urban gradients, the landscape-level processes that accompany urbanisation and underlie current theories of disease emergence could be captured.

2.6 Concluding remarks and future directions

In this review I consider the role that urbanization plays in the emergence of zoonoses, through exploring the ecological complexity of wildlife-livestock-human interfaces. In doing so I argue that interfaces should be considered a critical component of disease ecology in changing urban landscapes, and echo a body of recent literature calling for greater ecological sophistication in epidemiological theories of disease emergence [84–87]. The majority of epidemiological studies use foundational concepts to study a single, or small number of well characterized host species and pathogens when investigating transmission and connectivity within multi-host systems. Whilst this approach is well established, and useful in developing

Description	Examples	Proposed level of wildlife-livestock-human contact
Urban ecotonal interfaces and fragmentation of natural ecosystems (anthropogenically derived habitat edges)	Forest edge; agricultural edge; incursions for natural resource harvesting; urban wetlands	Increasing contact between humans, livestock and wildlife (both non-synanthropic and synanthropic species)
Evolving urban landscape – areas of informally planned resource provision	Informal refuse dumps; increasingly intensive farming and associated value chains (low biosecurity); backyard farming	High contact between humans, livestock and synanthropic wildlife that is largely unmanaged
Managed urban landscape – areas of formally planned resource provision	Sewage plants; established intensive farming and associated value chains (high biosecurity)	Controlled contact between humans and livestock Very little contact between wildlife, livestock and humans
Managed urban landscape – areas of recreational habitat suitable for wildlife	Parks and recreation facilities; gardens	Very few contacts between humans and livestock, and livestock and wildlife Increasing contact between humans and synanthropic wildlife

Table 2.1. A framework for wildlife-human-livestock interfaces in a developing city such as Nairobi. This has been adapted from a broader conceptual framework describing types of wildlife-livestock-human interface and their characteristics, developed by Jones et al. [16].

frameworks upon which the empirical characterization of a known host-pathogen system can be determined (through mechanistic models) and interventions planned (e.g. ref [34]), focus on a single species or pathogen might hinder the detection of pathogen emergence within a structurally complex system by overshadowing the evolutionary and transmission processes that precede this. As signaled by the emerging field of community disease ecology (reviewed by Johnson *et al.* [87]), new approaches are required to investigate disease emergence, that shift focus from the pathogen to understanding the processes underlying emergence [35]. In response, disease ecologists have moved towards adopting principles from community ecology; including metapopulation and network theory, trait-based approaches and a consideration of processes acting across biological scales [27,53,84,86–89]. The development of new modelling techniques will play a key role, and several frameworks have been suggested, that focus on integrating broad methodologies and cross-disciplinary collaborations to investigate causation in disease emergence [53,90,91]. Such methods will be key to unravelling the structural complexity of ecological communities at wildlife-livestock-human interfaces, and thus understanding how they function as epidemiological systems *prior* to disease emergence.

Whilst the focus of this review is on disease emergence, I would like to highlight the relevance of the frameworks discussed in combination with the broader concept of urban interfaces, for studying antimicrobial resistance (AMR). Currently considered urgent One Health issues, it is likely that the emergence of AMR and zoonotic pathogens in urban areas are underlined by a similar set of societal and ecological drivers [92]. Given the current rate of urbanization, and potential for associated changes in societal structure, food systems and natural ecosystems to expose human and animal populations to novel pathogens, I recommend an interdisciplinary approach to studying urban human-wildlife-livestock interfaces, with the following aims; *i).* establish characterizations for potential 'high risk' interfaces that exist along gradients of urbanization, and identify processes that have led to their formation, *ii).* describe biological organization and community ecology at these interfaces, conduct surveillance for priority zoonotic pathogens (i.e. those with 'emergent potential') across host taxa, and study the evolutionary processes underlying cross-species transmission where it is detected (see Box 2), *iii).* at interfaces where transmission risks are identified, develop appropriate interventions that can be used to reduce risk of transmission. Given their epidemiological significance, interfaces represent a critical point of control for the transmission of zoonoses. A detailed discussion of control measures is beyond the scope of this review, but interventions could be implemented at an interface (i.e. preventative action such as husbandry and 'behavioral' changes) or policy level (for a complete review see Gortazar *et al.* [93]). If, as I discuss in this review, pathogen dynamics at interfaces are characterized by dynamic changes in community structure driven by abiotic factors, emphasis should be focused on studying epidemiological connectivity (i.e. pathways and heterogeneity of transmission) and how this changes longitudinally with time. Such studies will be crucial in identifying the dynamic processes responsible for driving changes in community structure and thus pathogen dynamics at different interfaces over time.

Chapter 3

Methods

3 Methods

3.1 Acknowledgements

The work in this thesis has been conducted as part of the UrbanZoo project, a five-year project aimed at investigating routes of zoonotic pathogen emergence in the city of Nairobi, Kenya, through the study of microbial transmission across its urban landscape. As a large, multidisciplinary effort, many people have been involved in the design of this project, and collection and processing of data that has been used in the analyses contained within this thesis. The following list details my contributions towards collection and generation of the data, and acknowledges the hard work, contributions and affiliations of others who have been involved.

JH contribution:

- Designed protocols to collect samples from wildlife, livestock and the outside environment, and led a team of veterinarians to conduct wildlife sampling within each household compound.
- Designed data collection tools for wildlife sampling, and designed and managed database containing wildlife samples and their metadata
- Designed wildlife and ecological questions included in the household questionnaire.
- Assisted Titus Imboma with avian surveys.
- Undertook GIS analysis to characterise land use within each household compound.

Study design:
Prof. Eric Fèvre[1,2], principal investigator (epidemiology)
Prof. Mark Woolhouse[3], (epidemiology)
Prof. Jonathon Rushton[1], (economics and food systems)
Dr. Tim Robinson[4], (spatial analysis)
Prof. Sam Kariuki[5], (microbiology)
Prof. Julio Davila[6], (urban policy and international development)
Prof. Erastus Kangethe[7], (epidemiology)
Prof. Cecila Tacoli[8], (sociology)

Dr. Catherine Kyobutungi[9], (epidemiology)

Questionnaire design:
Dr. Judy Bettridge[1,2]

Fieldwork – training and supervision:
Dr. Judy Bettridge[1,2]
James Akoko[2]
Titus Imboma[10]
Dr. Bernard Agwanda[10]
Dr. Paul Webala[11]

Fieldwork – data collection:
Dr. Judy Bettridge[1,2]
Titus Imboma[10]
James Akoko[2]
Maurice Karani[2]
Patrick Muinde[2]
Yukiko Nakamura[12], Veterinary student
Fredrick Amanya[2]
Lorren Alumasa[2]
Dishon Muloi[2,3], PhD candidate (epidemiology)
Erin Furmaga[13], MSc student (epidemiology)
Titus Kaitho[14]
Victoria Carbonell[2,15]
Elin Öhgren[16], MSc student (epidemiology)
Kelvin Monyami[2]
Sandy Morton[17]

Logistic and administrative support:
Victoria Kyallo[2]

Laboratory work – supervision:
Dr. John Kiiru[5]
Dr. Judy Bettridge[1,2]
Dishon Muloi[2,3], PhD candidate (epidemiology)

Laboratory work – bacteriology:
Tom Ouko and the Kenya Medical Research Institute (KEMRI) laboratory team[5]
Nduhiu Gitahi and the University of Nairobi (UoN) laboratory team[7]

Laboratory work – DNA extraction:
Dr. Judy Bettridge[1,2]

Laboratory work – antimicrobial sensitivity testing:
Dishon Muloi[2,3], PhD candidate (epidemiology)
Elin Öhgren[16]

Database Management:
Dr. Judy Bettridge[1,2]
Absoloman Kihara[2]
Jason Rogena[2]

Bioinformatics:
Dr. Melissa Ward[18,19]
Dishon Muloi[2,3], PhD candidate (epidemiology)

1. Institute of Infection and Global Health, University of Liverpool, UK
2. International Livestock Research Institute, Nairobi, Kenya
3. Usher Institute of Population Health Sciences & Informatics, University of Edinburgh, Edinburgh, UK
4. Food and Agriculture Organization of the United Nations, Rome, Italy
5. Kenya Medical Research Institute, Nairobi, Kenya
6. The Bartlett Development Planning Unit, Faculty of the Built Environment, UCL, London, UK
7. University of Nairobi, Nairobi, Kenya
8. International Institute for Environment and Development, London, UK
9. African Population Health Research Centre, Nairobi, Kenya
10. National Museums of Kenya, Nairobi, Kenya
11. Maasai Mara University, Narok, Kenya
12. Faculty of Veterinary Medicine, Hokkaido University, Japan
13. Department of Epidemiology, Columbia University, New York, US
14. Veterinary Services Department, Kenya Wildlife Service, Nairobi, Kenya
15. ETH Zurich, Zurich, Switzerland
16. Uppsala University, Uppsala, Sweden
17. Peponi House Preparatory School, Nairobi, Kenya
18. Centre for Immunity, Infection and Evolution, University of Edinburgh, Edinburgh, UK
19. Nuffield Department of Clinical Medicine, University of Oxford, Oxford, UK

3.2 Introduction

This thesis focuses on the analysis of data that spans a range of spatial (city – sublocation – household) and ecological (landscape – host community – individual host) scales. To address this, I have taken an adaptive approach that utilises analytical methods from a range of disciplines including landscape epidemiology, community ecology, population genetics, and traditional epidemiology. As such, there is a great deal of variation in the methods applied to generating and analysing data between each chapter. To avoid repetition, this chapter provides a detailed overview of methodological elements that are common to chapters throughout the thesis, including study design, data collection in the field, laboratory procedures, certain analytical approaches, and the statistical frameworks that underpin them. Additional methods relating to generation and analysis of data that are unique to particular analyses, are described in detail in the relevant chapters.

3.3 Ethics Statement

Ethical approvals for human and animal sampling were obtained through the Institutional Research Ethics Committee and the Institutional Animal Care and Use Committee at the International Livestock Research Institute (ILRI), Nairobi. Permits for sampling of birds, rodents and bats were obtained through the National Museums of Kenya (NMK), and permits for sampling of non-human primates (NHP) and small carnivores were obtained through Kenya Wildlife Service (KWS).

3.4 Study Design

The UrbanZoo project, based in Nairobi, Kenya from 2012-2017, aimed to utilise a landscape genetics approach to understanding the movement and sharing of pathogens in a major developing city. A key component of this project, within which the research contained in this thesis is nested, was the ’99 household project’, which focused on informal livestock keeping practices in urban households as a route of zoonotic disease emergence in humans. As such, households were selected with the

aim of maximising the spatial distribution and diversity of livestock keeping practices across Nairobi, and were chosen to capture three main criteria: socio-economic diversity, population distribution and livestock keeping practices. Geospatial mapping data, generated as part of a technical report produced by Institut Français de Recherche en Afrique (IFRA), was used to identify 17 classes of residential neighbourhood in Nairobi based on physical landscape attributes, which were subsequently verified by 817 household questionnaires [92]. Each of the 17 classes of neighbourhood were then ranked by average income and reduced into seven wealth groups. Administrative sublocations were mapped onto each wealth group, identifying a total of 70 possible sublocations, for which dominant wealth groups were calculated by extracting the proportion of population belonging to each neighbourhood class within the sub-location boundaries (Table 3.1). A total of 33 sublocations were selected to be included in the study, with the number of sublocations belonging to each wealth group chosen proportionately to the population density and the variety of neighbourhood classes in each of the seven wealth groups. Final selection of individual sublocations was aimed at maximising areas with high livestock densities, whilst ensuring coverage of other neighbourhood classes and geographical spread (Figure 3.1).

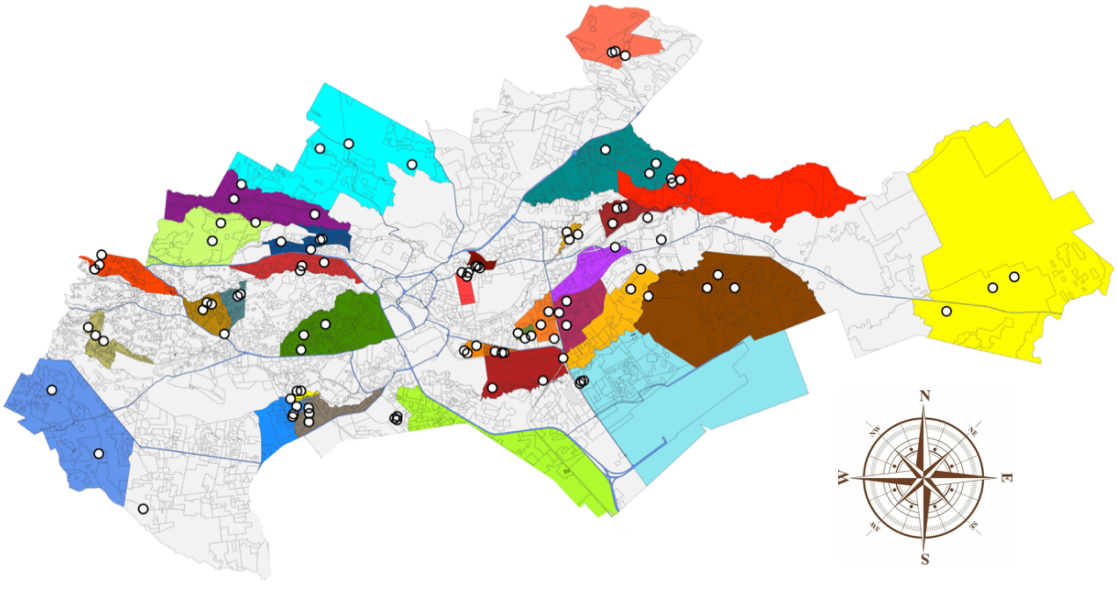


Figure 3.1. Map of the 100 households (white dots) and 33 sublocations (individually coloured areas) in Nairobi selected for inclusion in the study. Areas of the city in grey represent those not sampled.

Characteristics of physical neighbourhood classes identified by IFRA study (adapted from Ledant *et al.* [10])				Urban Zoo Project re-classification		
Tree cover	Defining characteristics	Neighbourhood description (housing type)	Average income	Wealth group	Possible sub-locations	Targeted sub-locations
> 13.5%	Detached housing with intense tree cover	Detached housing on very large plots *(>3000 m²)*	39,890	1	8	3
		Detached housing on large plots *(400 - 3000 m²)*	22,462	2	8	4
	Attached and semi-detached housing	Attached housing on medium plots *(<400 m²)* with important tree cover	22,084	2		
3% < 13.5%	Apartment building	Apartment buildings *with gated space*	22,084	2		
	Attached and semi-detached housing	Higher standing row houses *(plot size > 190 m²)*	13,352	3	5	3
		Lower standing row houses *(plot size < 190 m²)*	6,153	4	3	3
<3%	Roof cover >50% tiles	Lower standing *apartment buildings*	6,153	4		
		New areas of dense *single housing* development	3,855	5	9	5
	Roof cover > 40% concrete	High density multi-storey buildings	3,855	5		
3% < 13.5%	Apartment building	Apartment buildings *with open access*	3,855	5		
	Peripheral areas	Peripheral areas with residential component *(mainly residential)*	3,855	5		
		Peripheral areas with rural component *(presence of agriculture)*	2,165	6	24	11
	Collective housing	Community housing *with gated space*	2,165	6		
		Community housing *with open access*	2,165	6		
<3%	Roof cover >85% corrugated iron sheets	New areas of low quality housing *(built-up area <37%)*	2,165	6		
		High density planned low quality housing *(built-up area <37% AND public space >20%)*	2,165	6		
		High density unplanned low quality housing (slums) *(built-up area <37% AND public space <20%)*	1,301	7	13	4

Table 3.1. The seven wealth groups used by the Urban Zoo Project, and the number of sub-locations with a dominant wealth group identified and selected in the Nairobi municipality. Reprinted with permission from Bettridge *et al.* [2].

For each sublocation, three geographical points were selected at random within the dominant housing type. The order in which sublocations were visited was randomised. Local officials assisted in the recruitment of a household closest to each geographical point, to obtain two livestock keeping and one non-livestock keeping household per sublocation (a total of 99 households, 66 of which kept livestock). Households had to meet strict inclusion criteria of keeping either large ruminants (cattle), large monogastrics (pigs), small ruminants (goats/sheep), small monogastrics (poultry/rabbits), or no livestock species. To ensure an equal sample of both cattle and pig-keeping households, the combination of livestock keeping households represented in each sublocation was randomised, and had to consist of either large ruminant and small monogastric, or large monogastric and small ruminant species. For sublocations in which households keeping large ruminant or large monogastric species were absent, a replacement household keeping either small monogastic or small ruminant species was recruited. A single extra household was sampled in one sublocation, to replace a household in which the inhabitants did not consent to their faecal samples being collected and questionnaire data being conducted. As such, wildlife, livestock and environmental samples were collected from 100 households, and human samples and questionnaire data were collected from 99 households (Figure 3.1).

3.5 Data Collection

Sample collection

Two dedicated field teams were responsible for collecting data on humans, livestock and wildlife in each household, consisting of veterinarians, animal health technicians and clinicians. Informed consent was obtained from human participants, who were invited to submit a stool sample. Up to 20 rectal swabs were obtained from livestock species present in the household (ensuring that all species were represented in the sample). Rodents, bats, birds, small carnivores and non-human primates (NHPs) were all targeted for wildlife sampling. Rodents were trapped using medium-sized (23 cm x 7.5 cm x 9 cm) Sherman live traps (H. B. Sherman Traps Inc., Tallahassee, FL) or Victor lethal traps (Woodstream Corp., Lititz, PA) that were baited with dried fish, placed against walls throughout the household and livestock keeping facilities, and left

in place for three nights. Where possible, traps were set in each household for all trapping nights and checked daily. Mist nets were set at dawn to trap birds, with nets being positioned outside the house and around livestock keeping facilities. For household compounds in which bat activity was deemed likely (as judged based on the presence of fruiting trees and/or 'flyways'), mist nets were set at dusk and monitored for two hours. Where security conditions permitted, a remote bat detector (Song Meter ZC, Wildlife Acoustics, Inc.) was placed in each household compound for a single night, to monitor ultrasonic bat activity. If members of the household reported seeing small carnivores (such as mongoose) then Tomahawk cage traps (Tomahawk Live Trap Company, Tomahawk, Wis.) were set, baited with chicken and monitored regularly for a maximum of three days. Where NHP activity was reported at a household, wire-mesh live-capture traps were pre-baited with bananas for a minimum of three days. Traps were then set, and monitored regularly for a maximum of three days. Due to large variation in the size of household compounds, trapping effort (i.e. number of traps/mist nets placed per trapping session) was maintained such that it was proportional to the size of the household compound. The number of wildlife and livestock sampled are presented in Suppl. Table 3.1.

Once caught, all birds, and all but two bats caught per trapping session, were live-sampled in the field under manual restraint, before being released unharmed. Morphometric data was collected for identification purposes, and a suite of biological samples (including faeces if available, or a rectal/cloacal swab) were collected from each animal. All live rodents (except for individuals belonging to the genus *Cricetomys*, which were live-sampled under anaesthesia) and up to two bats caught per trapping session were transferred back to a biosafety level three (BSL3) laboratory, and humanely euthanised by cardiac puncture under isoflurane anaesthesia. Species identification was based on morphometric data. A full post-mortem examination was then performed, with fresh faeces being collected from the rectum. Rodents caught in lethal traps were also necropsied in the laboratory following the same protocols. Faecal samples were collected non-invasively from small carnivores, by keeping them in the trap for a maximum period of twelve hours. NHPs were anaesthetised where trapped, using a combination of Medetomidine and Ketamine (under the supervision of a Kenya Wildlife Service veterinary officer), and morphometric data and a suite of biological samples (including faeces if available, or a rectal swab) were collected from each animal. The primate was carefully monitored throughout, and anaesthesia reversed

using Atipamezol. Carnivores and NHPs were released unharmed at an appropriate time of day, from the same location at which they were trapped.

Outdoor environmental samples were collected from livestock pens and around the household compound using boot socks [93]. Water samples were collected from puddles and flowing water (i.e. rivers), using falcon tubes and modified Moore swabs respectively [94]. Moore swabs were formed from a tightly rolled piece of dishcloth folded horizontally in half several times and tightly rolled into a cylindrical shape. These were suspended in running water using fishing line for roughly 24 hours. Questionnaire and sample data was recorded using Open Data Kit (ODK) Collect software (Hartung et al., 2010), on electronic tablets, and uploaded to databases held on servers at the International Livestock Research Institute (ILRI) (for details of sampling forms, see Chapter 9, Tables A.1, A.2, A.3 and A.4).

Household and Individual Human Questionnaires

A nominated member of each household completed a questionnaire, detailing *i)* livestock ownership, management, sourcing, sales and antimicrobial use, and *ii)* household composition and socio-economic data. Abundance of livestock species and humans, and manure and household waste management practices were derived from this data for each household. Household composition and socio-economic data were used to generate 'wealth' and 'ruralness' indices for each household sampled [95]. These indices were calculated based on methods used to create the Demographic and Health Surveys (DHS) wealth index, which is derived from a Principal Component Analysis of easily measurable households assets (such as access to water, construction materials and ownership of livestock) [96]. A modification was made to the original set of household assets included in the DHS index to better capture household variation in Nairobi. An individual questionnaire was also administered to each human participant in the 99 household study, which captured demographic information, contact with livestock, perceptions of urban wildlife, and a detailed health assessment. For details of questionnaire data, see Chapter 9, Tables A.3 and A.5.

Ecological Surveys

Ecological surveys were used, alongside trapping data, to estimate the diversity of wildlife species present within households. Avian species counts (incidence) were conducted by a trained ornithologist from the National Museums of Kenya, in which species were identified based on audio-visual identification over a 20-minute period spent walking transects of each household compound. Surveys were conducted between 6:30am and 9:30am, over the course of two months in the dry season, ensuring that bird activity and weather conditions were constant. Where security conditions permitted, a remote bat detector (Song Meter ZC, Wildlife Acoustics, Inc.) was placed in each household compound for a single night, to monitor ultrasonic bat activity. As part of a household questionnaire, members of the household were also asked whether they saw rodents, small carnivores (such as mongoose) and NHPs within the household compound (with the aid of pictures). All species of wildlife documented in this study were grouped into functional groups, deemed relevant for the epidemiology of a directly transmitted gastrointestinal parasite such as *E. coli* (Table 3.2). Allocation of wildlife species to functional groups was based upon the

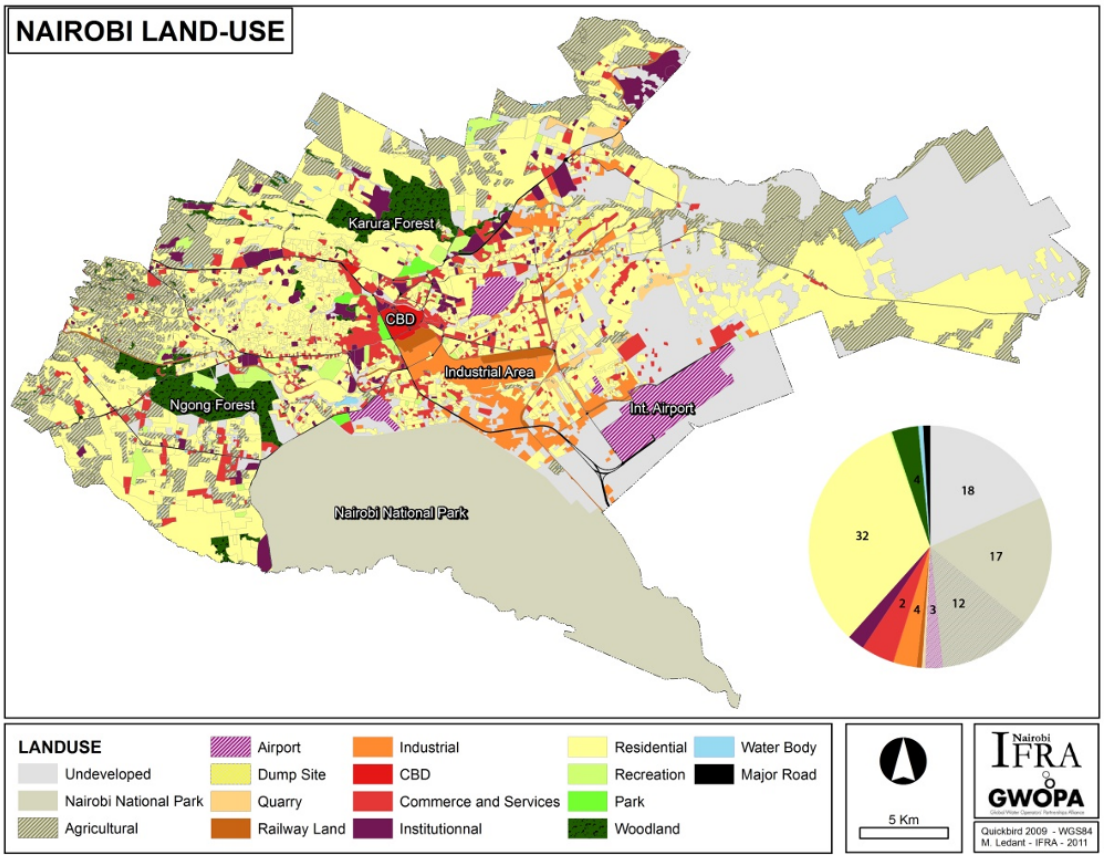


Figure 3.2. Map of Nairobi city, demonstrating different forms of land use across the city. Reprinted from Ledant [10].

EltonTraits database, as described in Wilman et al. [97]. These functional groups form the basis of species groupings for analyses in all subsequent chapters.

Land-use classification

Nairobi is characterised by a large variety of land use (Figure 3.2). The boundary of each household compound was drawn in ArcMap, and a 30m buffer created around the perimeter of each compound to represent the landscape surrounding it. Visual classification of land-use types within the compound and buffer area were conducted at 1:500 scale on a 1m resolution satellite-image available in ArcGIS. Characterisation of ecological characteristics along a perimeter around the household compound was considered as important, because the ecological setting within which the household exists extends beyond the boundaries of the compound. The extent to which this influential area of habitat outside the compound extends is unknown, and as such it was standardised across study sites. Within the boundary, the areas of eight different land-use types were visually identified and sketched as polygons; water-body, wetland, crops, mature trees, shrubs, grassland, bare ground, artificial ground and rubbish (descriptions for each of these are summarised in Table 3.3). The total area of classified land-use types at each site were calculated and expressed as proportions. Ecological land-use types (all except bare ground, artificial and rubbish) were used to calculate Simpson's α–diversity index, which considers both habitat richness, and an evenness of abundance among the land-use types present at each site. This index was created to represent the α-diversity of 'living' (biotic) habitat niches available to wildlife within households, and ranged from 1 (maximum heterogeneity) to 0 (absent). All classification was undertaken by the author who was familiar with the landscape at each site, and subsequently ground-truthed by revisiting sites.

Avian	Rodent	Bat	Primate	Carnivore
Plant/Seed-eating	Insectivorous	Insectivorous		
Omnivorous	Omnivore	Fruit		
Fruit/Nectar-eating				
Invertebrate-eating				
Vertebrate/Fish-eating /Scavenger				

Table 3.2: Functional groups by which wildlife species were grouped for most analyses in this thesis. Taxa were grouped into functional groups on the bases of their feeding ecology.

Land Use Class	Explanation
Water-body (environmental, biotic)	Natural water body (flowing or non-flowing)
Wetland (environmental, biotic)	Open wetland – reeds/rushes
Cropland (anthropogenic, biotic)	Row cops or other herbaceous crops (e.g. maize, coffee etc.)
Trees (environmental, biotic)	Trees (single or multiple), as determined by presence of a clear crown and evidence of shadow cast on ground (includes plantations of trees for commercial purposes [e.g. fruit])
Shrubs (environmental, biotic)	Shrubs, where able to distinguish from grassland and trees
Grassland (environmental, biotic)	Grass, pasture, herbaceous rangeland or bare-ground not serving a human purpose.
Bare ground (anthropogenic, abiotic)	Heavily compacted soil, serving human purpose (e.g. dirt road, playground)
Artificial (anthropogenic, abiotic)	Synthetic, man-made surface or object (including water-bodies) (e.g. tarmac road, cement, roof, swimming pool, water tank)
Rubbish (anthropogenic, abiotic)	Accumulation of human-derived waste

Table 3.3: Land use classifications, by which different forms of household land use were mapped

3.6 Bacteriology

All swabs and fresh faecal samples collected from wildlife, livestock and humans were placed in Amies transport media, and transported on ice to one of two laboratories (Kenya Medical Research Institute (KEMRI) or University of Nairobi (UoN)). Bootsocks and modified Moore swabs were transported in saline-filled polythene bags, on ice, and water samples were transported in falcon tubes on ice. Samples were enriched in buffered peptone water for 24 hours, and then plated onto eosin methylene blue agar (EMBA) and incubated for 24 hours at 37^{o}C. Subsequently, five colonies were selected and sub-cultured on EMBA, before being further sub-cultured on Müller-Hinton (MH) agar and stored at -20^{o}C in cryovials. A single colony was picked from each original sample (referred to as an 'isolate') and biochemical tests (sulphide indole motility test, triple sugar iron agar, urea agar and simmons citrate agar) were run for identification as *E. coli*. In keeping with the rest of the project, a single colony was picked from each avian/bat pooled faecal sample. Laboratory protocols were standardised between laboratories, and a post-doctoral research fellow was responsible for ensuring that these standards were maintained throughout the project. Further downstream processing of *E. coli* isolates (including antimicrobial sensitivity testing, DNA extraction in preparation for sequencing, whole genome sequencing (WGS) and bioinformatics) are explained in the relevant chapters.

3.7 Statistical Analysis

A brief introduction to data exploration, statistical and spatial models, and measures of ecological diversity that feature throughout this thesis is provided in this section. To avoid unnecessary repetition, other methods, where used, are described in the relevant chapters.

3.7.1 Data Exploration

A hypothesis-driven approach to data analysis was taken throughout this thesis, in which biological understanding of the underlying system was used to make *a priori* decisions about which models to fit. As such, data exploration was only used to test for and address potential statistical problems in the dataset that could inflate type I (rejecting the null hypothesis when it is true) and type II (failure to reject the null hypothesis when it is false) error. Following protocols described by Zuur *et al.* [98], the main sources of error tested for in the development of each model were the influence of outliers in the response or explanatory variables, collinearity between explanatory variables, and zero-inflation (the number of zeros in the data). Collinearity is one of the most important sources of error when determining the influence of a set of explanatory variables on one or more response variables, and becomes increasingly problematic when ecological signal is weak [98]. For each model, relationships between all sets of explanatory covariates were assessed using multi-panel pairwise scatterplots, Pearson correlation coefficients, and variance inflation factors (VIF). To reduce the influence of collinearity on model parameter estimates, variables with high VIFs were sequentially dropped, until all VIFs were smaller than a preselected threshold of 3 [98]. Only these variables were carried forward to the full model, and subsequent model selection procedures. Cleveland dotplots were used to identify outliers in response and explanatory variables. Outlying data points in response variables were addressed by selecting an appropriate probability distribution for use in the model (e.g. using a Poisson distribution for count data). Where outlying data points were present in explanatory variables (and were not believed to be true measurement errors), log base 10 or square-root transformations were applied to the variable in question.

Zero-inflation in response and explanatory variables was dealt with in several ways. Response or explanatory variables in which < 5% of data was present were considered data deficient, and therefore excluded from analyses. For models with a single response variable (e.g. generalised linear mixed effects models (GLMMs)), zero-inflation in the response variable was addressed by fitting a probability distribution that can account for this (e.g. zero inflated GLMMs). For multivariate ecological models, where response variables consist of the presence or abundance of multiple species across sites (either as raw data or a dissimilarity matrix), a high frequency of double zeros (i.e. where two species are absent from a site, for the same or different reasons) can complicate interpretation of analyses [98]. Households in this study were sampled over variable environmental conditions, and as such double zeros were common in the wildlife community dataset. To address this, asymmetrical dissimilarity coefficients (Hellinger transformation for tables of raw abundance or binary data, and Jaccard distance for binary data represented as a dissimilarity matrix) were applied to species response data, which replace differences in total abundance (or presence/absence for binary data) with relative variation in species composition between sites [99]. Data transformed in this way excludes double zeros, giving low weights to variables with many zeros.

3.7.2 Multivariate Canonical Models

Several approaches are available for testing hypotheses using community datasets, in which a table containing multiple response variables (e.g. species or genes) are regressed against a table containing a set of explanatory variables (such as environmental variables). Redundancy analysis (RDA) is used extensively in this thesis, and is a form of constrained ordination that combines the concepts of multiple linear regression and ordination to describe the variation in a set of response variables explained by a set of explanatory variables. Response variables $\mathbf{Y}$ (n x p) are regressed against explanatory variables $\mathbf{X}$ (n x m) which produces a matrix of fitted values $\ddot{\mathbf{Y}}$. A Principal Components Analysis (PCA) is performed on the fitted values $\ddot{\mathbf{Y}}$, which produces canonical eigenvectors and eigenvalues, and a matrix $\mathbf{Z}$ containing canonoical axes [100]. The most important canonical axes can then be plotted as ordination diagrams (triplots) that depict sites, response and explanatory variables. To address non-linear relationships introduced by sampling across an extensive

environmental gradient in Nairobi, transformation-based RDA (tb-RDA) was used, in which Hellinger transformation of the response data preserves Hellinger rather than Euclidean distances for RDA. Where hypotheses to be tested involve exploring variation in the dissimilarity (or 'distance') between objects (i.e. sites), a distance-based RDA (db-RDA) can be used. In this method, the response variable is represented as a distance matrix (in this thesis, distances between objects in the response variable were represented by the Jaccard distance coefficient), and input into a Principal Coordinate Analysis (PCoA). The resulting PCoA eigenvectors represent the dissimilarities in a Euclidean space and can be input as the response variables in a standard RDA, and regressed against the set of explanatory variables [101]. For certain multivariate data analysed in this thesis, the effects of a set of measurements **W** (e.g. spatial elements) could obscure interesting effects of the explanatory variables of interest **X** on the response variables **Y**. In these circumstances either a transformation- or distance-based partial-RDA was used, in which the effects of a **X** on **Y** were adjusted for the effect of **W** [102].

Model selection for all forms of global RDA model was performed using forward selection with a double stopping criterion, which aims to maximise the adjusted coefficient of multiple determination (adjusted R^2) at each step [103]. The double stopping criterion addresses two problems typically associated with forward selection methods (high type I error and including too many explanatory variables in the model) by running a global test on the full model first, and only progressing if that test is significant by permutation, and stopping the selection process if a candidate variable is deemed non-significant, or if it brings the adjusted R^2 of the model over the value of the adjusted R^2 of the global model [103]. To account for non-normal distributions in multivariate ecological data, significance testing was undertaken using non-parametric permutation tests. Permutation schemes (which describe how elements of the data are permuted in permutation tests) were structured according to the hypothesis being tested. For example, where hierarchical structuring of the data resulted in correlation between samples collected from the same household, this was accounted for by restricting permutations within household subgroups. All forms of canonoical analysis were computed in the R package "vegan" [104].

3.7.3 Generalised Linear Mixed Models

Mixed effects models are an extension of multiple linear regression that allow a dependency structure to be incorporated into the model. Including a random intercept that imposes a correlation structure on individuals drawn for the same group in the response variable, overcomes potential non-independence introduced by a hierarchical sampling design, which can lead to correlation of standard errors and inflated *p*-values [105]. Data in this study has two tiers of nestedness that are included as random effects in models throughout this thesis; data nested within households (i.e. samples) are grouped by household and sublocation, and household-level data is grouped into triplets by sublocation (Figure 3.3). At its simplest, a linear regression model with a single random intercept takes the form

$$y_{ij} = \beta_1 + \beta_2 \times x_{ij} + \alpha_i + \varepsilon_{ij} \quad \text{(eqn 1)}$$

where y_{ij} is the ith observation in group j, and α_i is a random intercept, which is assumed to be normally distributed with mean 0 and variance σ^2. The linear relationship between x and y (the fixed part of the model) is represented by a line with intercept β_1 and slope β_2. The intercept of the model is given by $\beta_1 + \alpha_i$, which changes randomly via α_i. Where necessary, the basic framework for a random intercept model described above was adapted for response variables with a discrete outcome (i.e. count or binary data), by applying models that replace the Gaussian distribution of linear regression with a non-linear Poisson distribution, Binomial distribution (with logit link function), or negative Binomial distribution, for count data, binary data for over-dispersed count data respectively.

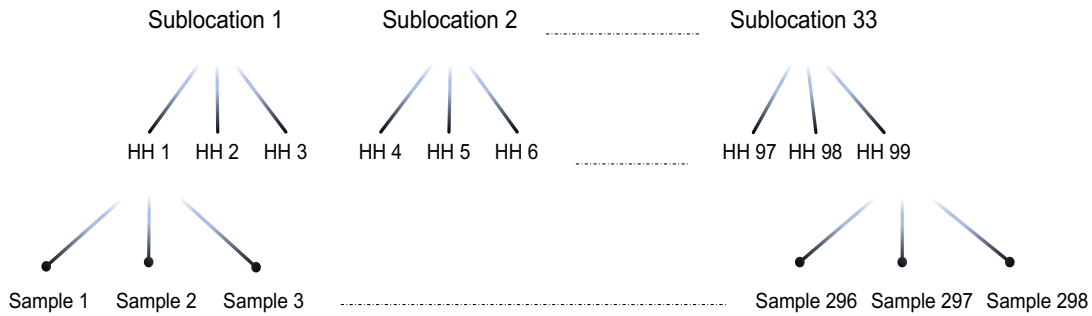


Figure 3.3: Hierarchical sampling design of the 99 households project. Samples are nested within households (HH), whilst triplets of households are nested within sublocations.

Throughout this thesis, models were constructed based on hypothesis-driven ecological reasoning, in which biological understanding of the system was used to build a global model by selecting variables *a-priori*. From global GLMMs, model selection was undertaken by comparing multiple, nested models $M_1,\ldots,M_N$ in a backwards, stepwise process using Akaike information criteria (AIC) [106]. AIC is a penalised log-likelihood that combines a measure of goodness of fit whilst requiring that the addition of parameters improves the log-likelihood of the model by a certain amount. An automated backwards selection procedure was used to select optimal GLMMs, starting with the full model and sequentially dropping the variable that gives the highest AIC value in combination until a model with the lowest AIC score was found (with the aim being to minimise AIC score). Likelihood ratio tests were also used to compare the fit of nested regression models. Model parameters for the final model were checked by plotting residuals versus fitted values, versus each covariate in the model. For Poisson and negative Binomial GLMMs with a dispersion parameter p, overdispersion was estimated using Pearson residuals and, if present, investigated further. Where necessary, Generalised Additive Models (GAM) were used to examine whether relationships between continuous variables could be modelled linearly. The residuals from each model were also assessed for spatial dependency, and plotted as a semivariogram. GLMMs were implemented in the R packages "lme4" [107] and "glmmADMB" [108], and GAMs were fitted in R package "mgcv" [109].

3.7.4 Modelling Spatial Structure and Generating Maps

Urban environments and their host/parasite communities are spatially heterogeneous, and understanding spatial structures present across the urban landscape and in these communities, could indicate the underlying processes that have created them. As such, it is important to capture this spatial structure in statistical models. In complex ecological and epidemiological systems, spatial structure operates across multiple scales, and a single response variable can display structure at more than one spatial scale [110]. To address this, a statistical method called Distance-based Moran's eigenvector maps (dbMEM) was used to represent spatial structure across all scales in the models in this thesis [111]. This approach begins with identifying the scales at which spatial structure (autocorrelation) is present in the response variable. The dbMEM base functions (eigenvectors) which represent structure in the response variable, are generated through a PCoA performed on a matrix of geographic distances between samples. Eigenvectors modelling positive spatial correlation are extracted, and regressed against the response variable to determine a set of significant dbMEM base functions modelling spatial structure in the response variable. These base functions can then be included as explanatory variables in the global model, or their variance can be removed from the model if spatial structure is deemed a 'nuisance variable'. To make them ecologically meaningful, dbMEM base functions can be split into arbitrary groups representing different spatial scales. Generally, base functions can be ordered from broad- to fine-scale as their number increases (e.g. dbMEM1-5 broad, dbMEM 6-14 medium, dbMEM 15-20 fine) [110]. dbMEM analyses are implemented in the package "adespatial" in R [112].

Maps of Nairobi were generated from shape files, and read into R as spatial polygon dataframes using the package "rgdal" [113]. Data was visualised and plotted onto these spatial vector objects in R.

3.7.5 Measures of Diversity in Community Ecology

In community ecology, the diversity of species communities can be measured in the three ways. α–diversity represents local diversity or the diversity of species at a site, and is represented as species richness (the raw number of species present) or by

an α–diversity index (such as Shannon or Simpson diversity). β–diversity represents variation in the structure of species communities between sites, within a region of interest, being defined by Whittaker [114] as "the extent of change in community composition, or degree of community differentiation, in relation to a complex gradient of environment, or pattern of environments". γ–diversity (or regional diversity) represents the species diversity in a region of interest. If β–diversity is constant across a region, then α– and γ–diversity are equivalent to one-another. Measures of α– and β–diversity are used extensively in this thesis to explore the response of host and bacterial genetic communities to urban land-use change. β–diversity between pairs of sites can be represented by a dissimilarity coefficient (such as the Jaccard index for binary data or Bray Curtis index for abundance data), which, across multiple sites, can be expressed as a dissimilarity matrix. In this form, hypothesis testing for determinants of variation in β–diversity between all sites across a region can be undertaken. Alternatively, a single value for β–diversity between all sites, BD_{Total} (which is equivalent to the total variance in a matrix of sites by species, **Y)**, can be calculated by dividing the total sum of squares in the dissimilarity matrix, SS_{Total}, by (n–1) [115]. When BD_{Total} is conceived as the total variation in **Y**, the contribution of individual sites to β–diversity can be calculated by dividing the diagonal values of a Gower-centred matrix of the centred dissimilarities (**G**) by SS_{Total}. Termed local contributions to β–diversity (LCBD), these values represent the uniqueness of each site in terms of community composition, and variation in the uniqueness of sites in terms of their community composition can be explored through hypothesis testing [115].

3.8 Glossary

The following list is intended as a quick reference for terminology, abbreviations and methods that feature in this thesis, and may be unfamiliar to the reader.

α–diversity: species diversity at a site

β–diversity: variation in species diversity between sites

Admixture (genetic): mixing (or interbreeding) of two or more genetic populations

Analysis of molecular variance (AMOVA): statistical method based upon analysis of variance, that is used to detect population differentiation utilizing genetic markers

Anthropogenic: pertaining to human impact on the environment (environmental change)

Antimicrobial resistance genes: genes conferring antimicrobial resistance

Asymptotic non-parametric models: statistical models (e.g. Chao2) that compare the estimated asymptotes of species accumulation curves, to estimate total species richness in communities. Non-parametric estimators are valid for all species abundance distributions

Core genome multilocus sequence typing (cgMLST): A whole genome approach to defining bacterial clones. Extends the number of genes included in classical multi-locus sequence typing (MLST) from seven to potentially >1000 genes, through a genome-wide gene-by-gene comparison approach

Degree centrality: A count of how many neighbours (or 'links') a node has in a network

Distance-based Moran's eigenvector map: A statistical method used to represent spatial structure in a dataset across all available scales

Epidemiological compartment: an epidemiologically functional host population, relating specifically to differentiation between the community of species on either side of an interface (i.e. human vs livestock vs wildlife)

Epidemiological connectivity: pertaining to the presence or absence of a pathway between two epidemiological compartments, across which parasites can be exchanged. Purposefully avoids inferring directionality of transmission

Epidemiological gradient: broad (i.e. regional)-scale gradients across which the structure of parasite communities vary, as a result of changes to their abiotic and/or biotic niche

Extensive drug resistance (XDR): bacterial non-susceptibility to at least one agent in all but two or less antimicrobial classes [26]

Hellinger transformation: an ecologically-motivated transformation that can be applied to binary and abundance community ecological data. This transformation is particularly suited to sites sampled over long environmental gradients, where species have low counts and many zeros

Isolation by distance (IBD): genetic populations are differentiated by limited dispersal of animals or parasites through space

Jaccard's Index/Coefficient: the Jaccard coefficient represents a binary measure of the ecological similarity and resemblance between two objects

Kendall's W coefficient of concordance: a non-parametric statistical test used to test the measure of agreement between quantitative variables and a set of objects that they are deemed to reflect. In this thesis, the statistic is used to test the concordance with which households are grouped into clusters by self-organising maps, based on characteristics of their wildlife, livestock and human communities.

Land-use change: changes to the structure of ecosystems as a result of human activities, which lead to perturbation of biotic systems. Examples include: deforestation, expansion of agriculture, pollution, depletion of marine fisheries, and eutrophication

Mobile genetic elements (MGEs): mobile elements of the prokaryote genome, which can be horizontally transferred between microorganisms and may confer adaptive functional traits such as virulence or antimicrobial resistance

Multi-drug resistance (MDR): bacterial non-susceptibility to at least one agent in three or more antimicrobial classes [26]

Mode: clusters of nodes (derived from self-organising maps and hierarchical clustering) with similar characteristics

Niche: abiotic ('physical') niches are used to describe a set of conditions that permit persistence of a host within a certain geographical range (such as climate), while the host species biotic niche is characterized by the inhibitory or facilitatory impact of other species (plant or animal) on its existence

Pan-drug resistance (PDR): bacterial non-susceptibility to all antimicrobial agents tested [26]

Pearson's phi coefficient of association: a measure of the degree of association between two binary variables

Sink/Source: terms used to represent the net flow of an agent (e.g. pathogen, AMR gene) between epidemiological compartments on a broad scale

Similarity profile routine: a series of permutation tests run on biological data that looks for statistically significant evidence of clustering in *a priori* unstructured sets of samples

Simpson's index (biotic habitat diversity): a measure of α–diversity, often used to measure the biodiversity of a habitat. It accounts for the number of species present, and their relative abundance. In this thesis it is used to represent the α–diversity of natural land-use types within household compounds

Species accumulation curves: community ecology tool used to estimate the species richness in a particular area. They record the cumulative number of species occurring in an area, as a function of the effort involved in sampling that area

Species accumulation curves: community ecology tool used to estimate the species richness in a particular area. They record the cumulative number of species occurring in an area, as a function of the effort involved in sampling that area

Spillover: the disease dynamics that enable a pathogen to be transmitted into a susceptible target host population from its reservoir population

Sublocation: administrative areas in Nairobi, used in

Transmission potential: used interchangeably with degree centrality. Reflects the potential for a node within an epidemiological network to spread parasites relative to other nodes in the network

Variance partitioning: an ecological method used to compare the variance explained by two or more sets of complementary hypotheses tested against the same ecological response variable

Virulence genes: genes conferring bacterial virulence

Weir and Cockerham's F_{ST}: the fixation index (F_{ST}) is a widely-used measure of genetic differentiation between biological populations. Weir and Cockerham's estimate of F_{ST} is unbiased with respect to sample size

Wildlife functional group: grouping of wildlife species by taxonomic and ecological traits that could inform their epidemiological role in parasite transmission

3.9 Supplementary Data

Taxonomic group	No. of individuals sampled	No. of species sampled (wildlife)
avian	547	57
avian populations	26 samples (4 pools)	4
bats	44	6
bat populations	20 samples (4 pools)	2
cattle	68	-
carnivores	5	2
chickens	271	-
ducks	27	-
geese	21	-
goats	112	-
g-pigs	4	-
g-fowl	21	-
pigs	53	-
pigeons	14	-
primates	4	2
rabbits	50	-
rodents	148	6
sheep	26	-
turkeys	10	-

Table 3.1: Number of individual wildlife and livestock sampled as part of the 99 households project

Chapter 4

Changes in Urban Land Use Drive Variation in the Structure of Wildlife-Livestock-Human Interfaces

Data used in the analyses contained within this chapter is available via an open access repository held by the University of Liverpool (datacat.liverpool.ac.uk/470).

4 Urban land-use change drives variation in the structure of wildlife-livestock-human interfaces

4.1 Abstract

Urbanisation can have profound impacts on the distributional ecology of hosts and epidemiology of infectious disease within them, but there is little empirical evidence that directly links variation in urban land use to the structure of host communities. The structure of household wildlife, livestock and human communities across Nairobi is related to data on household land use, demographics and socioeconomic status to explore how urban environmental change impacts communities of sympatric wildlife, livestock and human hosts. I describe variation in the composition of wildlife-livestock-human 'interfaces' across the city, and identify ecological and anthropogenic drivers correlated with the structure of wildlife and livestock populations. Furthermore, I demonstrate that host assemblages become less stable as the diversity of their habitat decreases. These results provide insight into the mechanisms by which host communities are restructured by urbanisation. An understanding of associations between urban drivers and host community structure is required to link urbanisation to the dynamics of microbial communities, and ultimately, human health outcomes.

4.2 Introduction

Urban development is an important form of land-use change, providing habitation and a source of livelihood for an increasingly large proportion of the world's population [3,116]. Rapid urbanisation is characteristic of the growth of developing cities, resulting in substantial changes to ecological systems and societal structure [2,117]. Subsequent changes to the structure and fitness of wildlife, livestock and human populations can be expected to generate novel interfaces across which new pathogens can emerge or existing pathogens can be maintained [9].

It is estimated that at least 70% of emerging zoonoses have a wildlife origin [11]. However, zoonotic transmission directly from wildlife is thought to be uncommon, and frequently relies on livestock species acting as a bridge for wildlife-human transmission [118]. Ecological change, and the resulting 'mosaic' of new and modified habitats created through urban development, generally lead to a decline in biodiversity, whilst presenting opportunities for wildlife species (such as rodents, birds and bats) that can co-exist with humans to thrive [23,119]. Such urban-adapted species ("synanthropes") may act as hosts for important bacterial, viral and parasitic pathogens, and mobile genetic elements conferring resistance to antimicrobials, shared with livestock [37,85,120,121] (reviewed in [122,123]). Recent studies have identified broad patterns in which synanthropic species are competent hosts for zoonotic pathogens [60,124,125]. However, the role of synanthropy as a trait that influences the evolutionary behaviour of multi-host pathogens has yet to be investigated.

Considerable variation exists in how wildlife use the urban landscape for activities such as foraging or reproduction [126]. It is this variation in the utilisation of resources made available to them in such environments, in conjunction with the nature of these resources (source, spatial distribution etc.), that will determine interactions between wildlife and livestock, and the subsequent establishment of wildlife-livestock interfaces [9,90]. Urban agriculture forms one such resource, and is increasingly practiced in developing countries as a result of growing demand for animal-sourced food products and lack of traditional supply-chain infrastructure, and as a form of subsistence for the urban poor [127,128]. Land shortages, particularly in informal settlements where population growth and density are highest, mean that livestock are commonly kept within household compounds, or left to scavenge freely

in communal areas [129]. Such "backyard" management is often characterised by low biosecurity, mixed-species farming, feed sourced from the household or local area, and poor hygiene and management of manure disposal, which taken together can promote wildlife-livestock contact and generate favourable conditions for disease transmission [87,130]. Changes in wealth and land use along gradients of urban development result in varied urban livestock keeping and management practices, which would be expected to generate diverse wildlife-livestock interfaces across livestock keeping households [95]. In considering how urbanisation could drive emergence of directly transmitted zoonotic agents through changes to the distributional ecology of hosts, it is helpful to consider urban land-use change as a combination of abiotic ('physical') factors that can influence the biotic ('living') and abiotic niches within which host and pathogen species exist. The theory behind ecological niches has existed for more than a century, and over that period niches have been variably used to describe how resource provision and interspecific competition influence species' distribution and abundance. More recently, niche-based thinking has been successfully applied to model species responses to environmental change [131]. For the purposes of this study, the terms environmental and anthropogenic are used to make a clear distinction between 'natural' and 'unnatural' processes and niches.

The process by which disease spillover occurs at an interface can be thought of at two scales; i) the ecological influence of landscape changes on the distribution and density of reservoir and target hosts (and thus the 'pathogen pressure' on target hosts), and ii) a series of epidemiological and behavioural limiting steps (or 'barriers') to transmission between reservoir and target hosts, and replication in and dissemination amongst target hosts (reviewed by Plowright et al. [132]). An expansive body of theory exists to describe processes acting at the former scale, but there is little empirical evidence that directly links changes in the function of abiotic and biotic systems to the structure of host communities. Understanding the form that host communities take in cities is necessary to link urban environmental change with human health outcomes, and inform disease surveillance efforts. In this chapter, methods from community ecology are used to describe the form of wildlife-livestock-human interfaces across urban households in Nairobi, and investigate how variation in urban land use influences their host community structure. If urban development is responsible for structuring assemblages of wildlife, livestock and human hosts, I expect to identify groups of similarly structured interfaces across the urban landscape,

the form of which are determined by the availability of a set of biotic and abiotic niches (for wildlife), and anthropogenic determinants (for livestock). As such, my approach addresses three questions: i) what are the city-wide characteristics of household land use (which I consider an important component of the biotic and abiotic niches within which host communities exist) and the broad and fine-scale structures of wildlife, livestock and human populations (the biotic niche of the microbial community), ii) can variation in land use and host characteristics be used to group households with comparable ecological structures, thus identifying categories of similarly structured urban wildlife-livestock-human interface, and iii) how does variation in household land use associate with the structure of wildlife and livestock host communities.

Finally, to explore the epidemiological interface existing between synanthropic wildlife and livestock at a global scale, a database of 2073 host-zoonotic parasite associations (representing 436 species of mammalian and avian hosts, and 504 zoonotic parasites) is created. This is used to generate and compare quantitative measures of zoonotic parasite carriage for similar groups of wildlife and livestock hosts to those sampled in Nairobi.

4.3 Methods

The urban classification of wildlife-livestock-human interfaces described by Hassell et al. [9] was used to define interfaces of interest in this study. Household livestock keeping was focused on as representing points of largely unmanaged, intense contact between synanthropic wildlife and livestock.

4.3.1 Data collection

Procedures for selecting household interfaces are described in detail in Chapter 2.4. Land use comprises the biotic and abiotic niches within which hosts exist, and was classified for each household as previously described (Chapter 3.5, Table 3.3). Data was collected on the presence of avian species, rodents, fruit bats, insectivorous bats, non-human primates (NHPs) and small carnivores in each household compound. These wildlife species were grouped into epidemiologically relevant functional

groups, that took into consideration both their feeding and positional ecology (Table 4.1), and the incidence of each functional group was recorded in each household. The presence of each functional group was estimated using a variety of methods, combining biological sampling with ecological surveys and the household questionnaire, details of which are described in Chapter 3.5. For birds, abundance of species per functional group was also calculated in each household. Allocation of wildlife species to functional groups was based upon the EltonTraits database, as described in Wilman et al. [97]. The incidence of wildlife species/functional groups and livestock species within each household was used to estimate α-diversity (richness) of these communities. α-diversity of wildlife was determined by combining avian species richness (the total number of avian species recorded in a household) with the incidence of mammalian functional groups in each household.

Avian Functional Groups		Rodent Functional Groups	Bat Functional Groups
Feeding Ecology	**Strata**	**Feeding Ecology**	**Feeding Ecology**
Plant/Seed-eating	Low Canopy	Omnivore	Insectivorous
Omnivorous	High Canopy		Fruit
Fruit/Nectar-eating			
Invertebrate-eating			
Vertebrate/Fish-eating /Scavenger			

Table 4.1: Epidemiologically relevant functional groups by which wildlife species were grouped, accounting for both feeding and positional ecology

Livestock and human communities, and anthropogenic determinants.

Abundance of livestock species and humans, manure and household waste management practices, and wealth and ruralness indices for each household were derived from questionnaire data (for further details see Chapter 3.5). Dividing livestock and human abundance by household area (m^2, as measured using ArcGIS) generated an estimate of density of livestock and humans in each household. Each human participant in the study completed a separate questionnaire, detailing their health and any occupational exposure they had with livestock or their products. Different forms of interaction with livestock were recorded (including herding, direct handling, milking, egg-collection, slaughter, and contact with housing, and manure) and each individual was ranked according to their perceived exposure to livestock-borne zoonoses through these activities (from 1-3). This index was averaged across all humans within each household, to generate a household measure of livestock contact.

Health parameters (age, height and weight) were used to calculate an average body-mass index (BMI) for each household. Following World Health Organisation (WHO) guidelines, BMI cut-offs for underweight adults over 20 years of age were set at <18.5. Underweight children between the ages of 5 and 19 were identified as having a BMI-value-for-age < -2 SD (based upon Z-scores of BMI) [133].

Host-parasite Association Database

To generate indicators for zoonotic parasite richness and sharing between wildlife and livestock hosts, a global database of species-level mammal-parasite and avian-parasite interactions was created by combining data from two existing databases. The ENHanCEd Infectious Diseases (EID2) database details 22,515 unique host-parasite associations between 6,314 parasites (bacteria, viruses, fungi, vectors and prions) and 8,905 hosts (mammal, reptile, fish, amphibian, bird and plant), and is sourced from published literature and sequences [134]. The Olival database details 2,805 mammal-virus associations for 586 unique viral species found in 754 wild and domestic mammalian species, and is sourced from published literature [135]. Combined, these databases represent the most complete set of host-parasite interactions available. For the purposes of this exercise only directly transmitted zoonotic parasites were considered, and as such vectors and prions were removed from the database. Host-parasite associations were restricted to mammalian and avian host species exhibiting synanthropic traits, and hosts that utilise human-altered landscapes were identified and filtered for by querying the IUCN Red List for avian and mammal species with habitat listed as "Artificial/Terrestrial". This search criteria returned species with habitat preferences listed as "arable land", "pastureland", "plantations", "rural gardens", "urban areas" or "subtropical/tropical heavily degraded former forest". Species taxonomy was standardised between the EID2 and Olival databases using the mammalian taxonomic database Mammal Species of the World, and the avian ecological database EltonTraits 1.0 (which follows avian taxonomy as described in the BirdLife V3 world list) [97]. Domestic species represented in the UrbanZoo study (cows, goats, sheep, pigs, chickens, turkeys, guineafowl, ducks, geese and lagomorphs) were also retained in the database. For the purposes of this study, the working definition of a zoonosis included any parasite sharing both human and animal associations in the database. As such, only host-parasite associations for parasites which also maintained a human-parasite association were included. Wildlife hosts

were grouped into the same functional groups (without the strata classifications) considered in the UrbanZoo project (Chapter 3.5, Table 3.2).

4.3.2 Data Analysis: Describing the structural form of household wildlife-livestock-human communities, and identifying groups of households with similarly structured host communities.

My analytical approach is depicted in full in Suppl. Figure 4.1. All analyses were conducted in R [136]. I began by describing the ecological structure of host and pathogen niches within households sampled across the city. Self-organising maps (SOMs), a form of unsupervised machine learning conducted within a framework of artificial neural network algorithms, were used to describe variation in land use and broad patterns of α–diversity and density in wildlife, livestock and humans across households. SOMs behave in a similar way to clustering algorithms, providing a discretised representation of variables in space, whilst preserving the topographical structure of the input data on a map [137]. The map portion of a SOM (the 'component planes') is particularly useful for visual interpretation of relationships within high-dimensional multivariate datasets. Consequently, SOMs represent a valuable addition to the ecologists toolbox, when dealing with complex, non-linear relationships in multivariate datasets [138]. Two SOMs were constructed in the R package 'kohonen' [139]; one based on the proportion of different land-use types, and the other on α–diversity of wildlife and livestock, and density of livestock and humans within households. The training dataset was scaled and centred, before being run with 1000 iterations and a learning rate from 0.05 to 0.01. Diagnostic quality of the model was checked by evaluating training progress and node counts. The final map consisted of 7x8 nodes in a hexagonal configuration, with each node representing an array of values corresponding to the input variables. Clusters of nodes with similar characteristics (henceforth termed household 'modes' and/or 'modal groups') were identified using hierarchical clustering, with estimates for a reasonable number of household clusters ascertained using a k-means algorithm and examining a scree plot of within-clusters sum of squares. A similarity profile routine (SIMPROF) test was also used to evaluate modal clustering of nodes, in the R package 'clustsig' [140]. Kendall's *W* coefficient of concordance, tested by permutation and Holm corrected for multiple testing, was applied to the results of the SOM (which consisted of 56 nodes, [each node

representing a prototype household], and their modal groupings), to test whether clustering results were globally significant. To identify the contribution of each input land-use category to variance between SOM nodes (and thus clustering/grouping of the data), a Bayesian approach to feature significance was used, in which the probability of each feature (i.e. variable – in this case land-use category) capturing the structure of the data was compared within a probabilistic framework in the R package 'popsom' [141].

Fine scale patterns of wildlife and livestock co-occurrence within households were explored by applying a combination of a SOM and statistical analysis rooted in ecological theory to presence/absence data for wildlife functional groups and livestock species within households. A SOM was developed to represent co-occurrence of all wildlife taxonomic groups and livestock species. The variables livestock density, human density, human health and livestock contact were included in the model to visualise relationships between host assemblage and these broader characteristics of the host populations at interfaces. To relate the community structure of households to their broader scale modes, presence/absence data was Hellinger transformed, generating a quantitative measure of the relative contribution to household wildlife diversity (richness) for each species present in a household [99]. Associations between these species' weightings and household modes were assessed using Pearson's *phi coefficient of association* statistic, and tested for significance using 10000 permutations [142]. Further details of the statistical tests used in this section are provided in the methods chapter.

4.3.3 Data Analysis: Associations between urban land use and host community assemblages at household interfaces.

To determine how variation associated with urban land use affects the structure of wildlife and livestock host communities, and thus the biotic niche available to pathogens within households, statistical models were used to test four hypotheses. The first two hypotheses considered determinants for the finer-scale structure of household wildlife and livestock communities separately, testing whether *i)* wildlife community structure varies according to a set of environmental (biotic) and/or anthropogenic (abiotic) niches, related to resource provision within households, and whether *ii)* the structure of livestock communities varies by social determinants such as wealth. The

final two hypotheses considered drivers of dissimilarity (β–diversity) in wildlife-livestock community composition between households, testing whether *iii)* the structure of wildlife-livestock host assemblages was associated with differences in environmental and/or anthropogenic determinants, and whether *iv)* the compositional 'stability' of wildlife-livestock host communities in relation to one-another in urban settings is associated with the complexity of the biotic niches within which they exist.

Five response variables were represented in these models. Household wildlife and livestock assemblages were split into four community datasets; one with binary presence/absence of all wildlife functional groups per household (n=99), one with abundance of avian groups per household (n=99), one with abundance of livestock per household (n=66), and one combining presence/absence of wildlife functional groups with abundance of livestock per household (n=66). A Hellinger transformation was applied to the first three of these community datasets, to account for the high proportion of double-zeros in each dataset [99] (for further details see Chapter 3.7.1). The wildlife-livestock dataset was transformed into a distance matrix, with the Jaccard dissimilarity index representing dissimilarity (β–diversity) in wildlife-livestock community composition between households. This index is used in community ecology to measure the response of communities to gradients of ecological change [143]. Finally, local contributions to β-diversity (LCBD indices, derived by decomposing the total β-diversity represented in the community dataset (BD_{total}) into site and species-based contributions [115]), were used as a measure of the compositional 'stability' of wildlife-livestock host communities in relation to one-another. LCBD values are strictly positive, and increase as the community of hosts at each site becomes more unique; sites with large LCBD values could therefore represent urban ecological conditions in which host community structures depart from normality [144]. An LCBD value was calculated for each livestock keeping household (n=66) using the function *betadiv* in the R package 'adespatial' [112]. Spatial structure in each response dataset was represented using distance-based Moran's eigenvector maps (dbMEMs), which provide a powerful multivariate approach to model spatial structure in a response variable, and can be partitioned by broad, medium and fine spatial scales [111]. dbMEM eigenvectors modelling significant spatial variation were included as partial terms, thus removing spatial variation from the model.

For ‘wildlife’ models (hypothesis *i*), the influence of a detailed set of household-level covariates on variation in wildlife and avian community datasets was tested by partitioning the variance of environmental and anthropogenic determinants within a canonical redundancy analysis (RDA, details of all models are provided in Table 4.3, see models $W^1/W^2/A^1/A^2$). This approach permits partitioning of community variation among different sets of explanatory variables allowing variation to be split by anthropogenic variables [a|b], environmental variables [b|c], both environmental and anthropogenic variables [b], and an unexplained component [d] (Suppl. Figure 4.2) [102]. Statistical significance of each fraction with respect to all others was tested using RDA and analysis of variance (ANOVA). Correlation between tree cover and artificial land use resulted in specification of two candidate models for each response dataset. For the ‘livestock’ model (hypothesis *ii*), the livestock community dataset was regressed against wealth and ruralness indices, household area and biotic habitat diversity of households (considered important determinants for livestock-keeping practices) in a partial-RDA, permitting the presence of significant spatial structure in the response variable (represented as dbMEM eigenvectors) to be controlled for. All forms of canonical analysis were computed in the R package “vegan” [104].

A different set of environmental and anthropogenic variables were selected to represent possible determinants for both wildlife and livestock community structure in households (Table 4.3, models WL^1/WL^2, hypothesis *iii*). These variables were regressed against the wildlife-livestock community distance matrix in one of two partial-distance-based redundancy analyses (partial-dbRDA, specifying either tree cover or artificial land use). This approach conducts a principal coordinates analysis (PCoA) of the response variable, generating a set of principal coordinates that are subsequently used as response variables in a partial-RDA [101]. Effects of spatial structure in the distance matrix were represented and controlled for using dbMEM’s as a partial variable. To examine urban drivers of host community stability (hypothesis *iv*), household LCBD values were regressed against the same set of environmental and anthropogenic determinants described in models WL^1/WL^2, in two linear mixed effects models (LMM, Table 4.3, models $LCBD^1/LCB^2$, specifying either tree cover or artificial land use) in the R package ‘lme4’ [107].

In all models, household dependency in the sampling design was accounted for by either constraining permutations within sublocations (canonical models) or

including sublocation as a random effect (LMMs). Optimal canonical models were chosen using forward selection with double stopping criteria, implemented with the *ordiR2step* function in R package 'vegan' [145]. This approach only performs forward selection when the global model is significant, avoiding inflation of overall Type 1 error. Optimal LMMs were chosen using stepwise, backwards elimination from the full model based upon Akaike information criteria (AIC). Significance of model terms were tested by 999 permutations or maximum likelihood test for canonical and LMMs respectively, and the fit of each model was reported as regression coefficients of multiple determination (R^2_{adj}) for canonical models or marginal R^2 for LMMs. All canonical and LMM models (full candidates and their optimal derivatives) are depicted in Table 2. Data exploration and model validation procedures were carried out as described by Zuur *et al.* (2010).

4.3.4 Data Analysis: Estimates of zoonotic parasite richness and sharing in wildlife hosts with synanthropic traits, and livestock

Metrics of parasite richness and sharing were estimated from the host-parasite association database and used to explore the epidemiological interface between wildlife functional groups and livestock at a global scale. Host parasite richness was calculated as the total number of unique zoonotic parasites detected within each wildlife functional group or livestock group. Ecological networks, where nodes represent species, connected by sharing of parasites, provide an intuitive way to represent parasite sharing between host groups [146]. For host-parasite sharing, 'degree centrality' [a measure of network centrality that represents the total number of parasite associations each node (host group) shares with other nodes (host groups)] was used to represent the breadth of parasite-sharing between wildlife and livestock groups. Following Pilosof et al. [147], degree centrality was considered to reflect the potential of a node (host group) to spread parasites relative to other nodes in the network, and thus the term degree centrality is used interchangeably with 'transmission potential'.

Methods from community ecology were used to address bias introduced to the estimates of parasite richness and host-sharing through variation in research effort.

Species accumulation curves (SACs) and asymptotic non-parametric models have been used to address sampling bias in the context of ecological network analysis, by estimating the expected number of species or interactions as a function of the number of observations [148]. Separate SACs were fitted for each functional group, with host species comprising each functional group treated as observations. Estimates for parasite richness and degree centrality were derived using Chao2 non-parametric models, which generated an estimate for the asymptotic point where increasing the number of host species no longer increased the number of parasites or host-host associations for each functional group. SACs and Choa2 asymptotic models were fitted using the R package "Fossil" [149]. Ecological estimates of parasite richness and sharing for wildlife functional groups and livestock were visualised as an undirected, unipartite graph in the network analysis software Gephi, using the force-directed algorithm "OpenOrd" [150]. Raw species-parasite data from the database (un-corrected for research effort) was used to examine zoonotic parasite sharing between species belonging to wildlife and livestock functional groups. Associations were visualised as an undirected unipartite graph in the network analysis software Cytoscape, using the force-directed algorithm "Edge-weighted Spring-Embedded Layout" [151]. The H-index (which has been used as a measure of scientific interest in a disease [152]) was adapted to represent research effort for each functional group. A functional group has index h, if h of the Np host-parasite associations have at least h citations each, and the other Np-h parasites have no more than h citations each.

4.4 Results

4.4.1 Structural form of wildlife-livestock-human interfaces across households in Nairobi

Six land-use categories; artificial, bare ground, grassland, crops, shrubs and tree cover (representing 49.4%, 14.9%, 3%, 12.9%, 6.7% and 12.5% of land cover respectively), were widespread in households across Nairobi, whilst waterbodies, rubbish and wetlands were each found in less than 10% of households. As such, only the former were used to determine habitat groupings. The optimal SOM reached a minimum plateau for the distance from each node's weights to the samples represented by that node at 800 iterations (Suppl. Figure 4.3). Inspection of component planes of the SOM indicated that tree cover, artificial and bare ground were responsible for most differentiation between prototype households, as supported by Bayesian feature significance estimates from the model (Figure 4.1). Using the within-cluster sum of squares statistic, 7 'habitat niches' were identified, which are based upon the distribution of SOM training variables (land-use types), and would be expected to support different wildlife-host communities (Figure 4.1). Although SIMPROF identified 8 significant hierarchical clusters, a more conservative selection was made to preserve sample size across habitats. Permutation tests applied to Kendall's W coefficient of concordance showed that household membership to all groups was statistically significant ($P<0.01$).

At a broad scale, variation between host community structure arose mainly from wildlife α-diversity (Bayesian feature significance: wildlife diversity, 0.683; livestock diversity, 0.305; livestock density, 0.008, human density, 0.004), whilst household densities of humans and livestock were highly correlated (Figure 4.2). Within-cluster sum of squares showed that households could be broadly separated into four groups, each of which displayed distinct trends in wildlife-livestock-human community characteristics (Table 4.2, Figure 4.2). These groupings were supported by SIMPROF analysis, which identified four significant groups, and permutation tests applied to Kendall's W coefficient of concordance which showed that household membership to groups 1, 3 and 4 was statistically significant ($P<0.001$). Household groups were spatially orientated across the city, with groups 1 and 4 being clustered to the West, group 2 centrally, and group 3 centrally and Eastern (Figure 4.2).

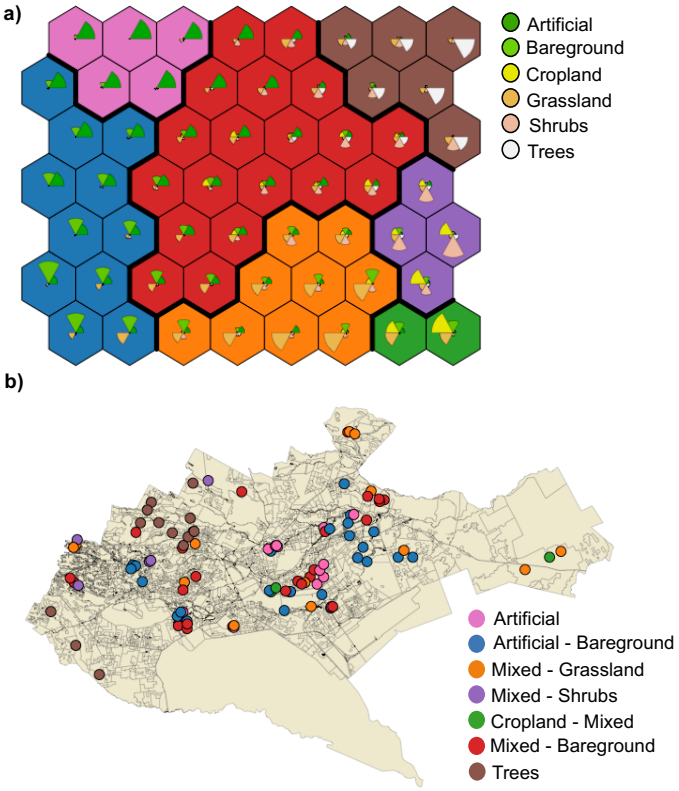


Figure 4.1: Self-organising map (SOM) and principal components analysis as applied to variation in land-use change in Nairobi. **a)** Households with similar land use characteristics are represented by SOM nodes (hexagons), which are coloured by the results of hierarchical cluster analysis ('habitat niche'). Pie charts within nodes (SOM node weightings) demonstrate the importance of each land-use category for that node. **b)** Households mapped by 'habitat niche' across Nairobi.

Household Mode	Human density	Livestock density	Livestock diversity	Wildlife diversity	Wildife associations	Livestock associations
1	High	High	Low - Medium	Low	Scavenging Birds, Rodents, Seedeating Birds, Insect Bats	Poultry (indigenous) Small Ruminant Lagomorphs Poultry (other)
2	Low - Medium	Low - Medium	Low - Medium	Low - Medium	Invertebrate (low) Omnivore (high) Omnivore (low) Seedeating Birds, Insect Bats	NA
3	Low	Low	Medium - High	Medium - High	Frui/Nectar (high) Invertebrate (high) Invertebrate (low) Omnivore (high) Omnivore (low) Scavenging Birds	Small Ruminant Lagomorphs Poultry (other)
4	Low	Low	Low - Medium	High	Frui/Nectar (high) Invertebrate (high) Invertebrate (low) Omnivore (high) Omnivore (low) Fruit bat Primate Carnivore Scavenging Birds	NA

Table 4.2: Characteristics of households as per their host assemblages. Households are grouped into 'modes' by hierarchical clustering performed on a Self-organising map (SOM) (columns 1:4). Fine-scale, statistically significant species associations with household modes (tested using Pearson's phi coefficient of association) are depicted in columns 5 and 6.

A SOM detailing finer scale, species-level structure of household interfaces showed clear patterns of distributional overlap within and between wildlife and livestock, enabling identification of sets of frequently co-occurring taxa, which could be separated into "generalists" (those widely distributed across households, and frequently co-occurring), and "specialists" (wildlife constrained to ecological niches, or livestock kept according to anthropogenic determinants) (Figure 4.3, panels 1 - 3). Trends between wildlife/livestock host group assemblage and livestock density, human density, human health or human contact with livestock were absent (Figure 4.3, panel 4). When linked back to broader scale household modes, there was clear differentiation in Hellinger-transformed species assemblages that were significantly associated with household modes (Table 4.2). No differences in livestock contact or human health were detected between modal groups (Kruskal-Wallis chi-squared: 1.6051, df=3, P=0.66; Fisher's exact: P=1).

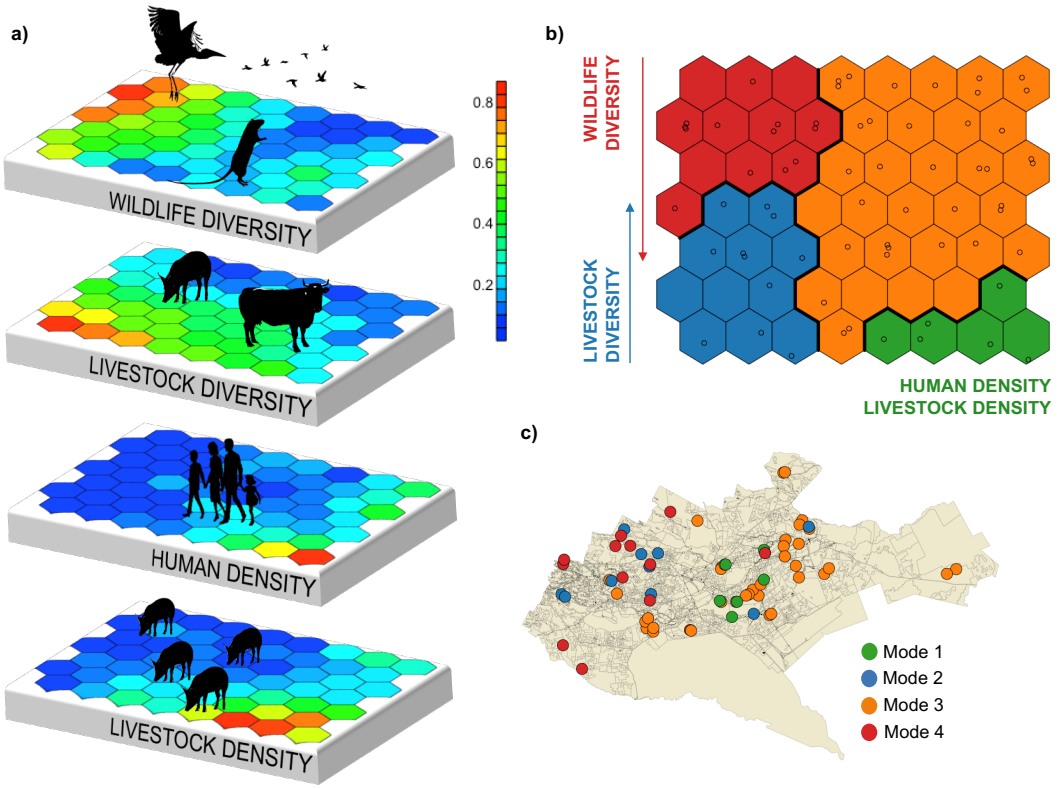


Figure 4.2: Self-organising map (SOM) as applied to broad scale community characteristics in households. **a)** Each input variable for the SOM is mapped onto a component plane, permitting visualisation of the magnitude of each variable in relation to the other variables across households. The colour scale (blue-red) represents the codebook vector value assigned to each SOM node. This is a vector of sample properties, in this case representing α-diversity of wildlife or livestock, human density or livestock density. **b)** Membership of households to four household 'modes', coloured by the results of hierarchical cluster analysis. Small circles represent households within nodes. **c)** Households mapped by modal membership across Nairobi.

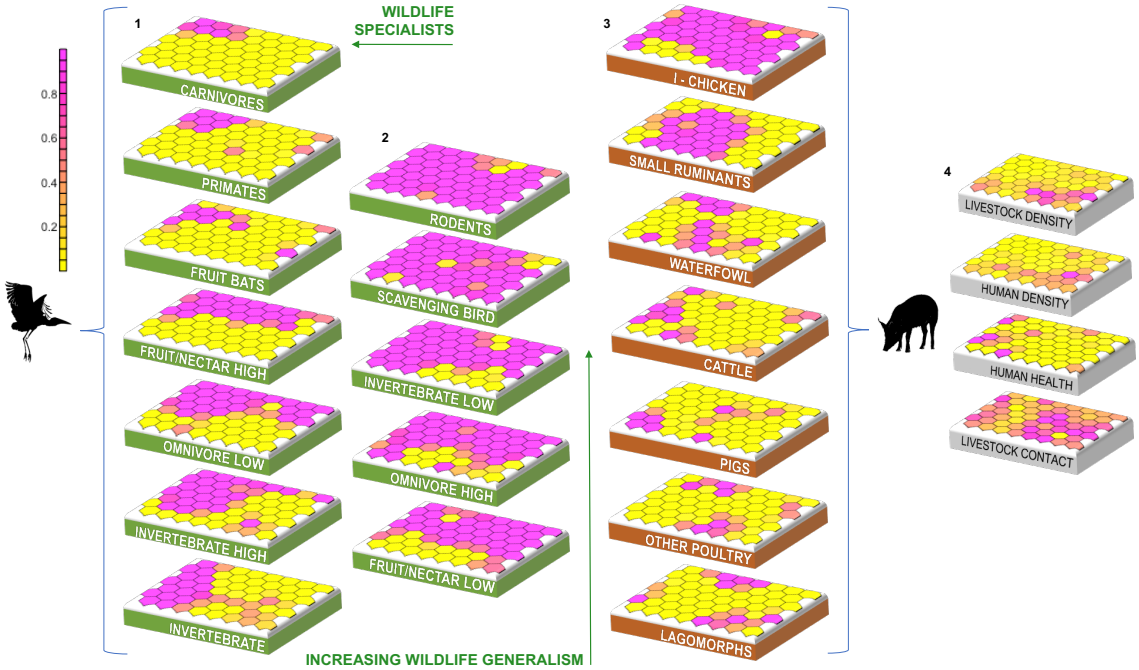


Figure 4.3: Self-organised map (SOM) as applied to fine scale community structure in households. **a)** SOM component planes for each variable [wildlife functional groups (green planes; panel 1 = 'specialist' wildlife functional groups, panel 2 = 'generalist' wildlife functional groups), livestock species (brown planes, panel 3), and disease determinants (grey planes, panel 4)].

4.4.2 Associations between urban land use and host population community structure in households across Nairobi.

Wildlife. The structure of wildlife communities was predominantly determined by land use within households, and there was no significant variation in wildlife or avian community structure between livestock and non-livestock keeping households [$F(1,97) = 1.102$, $p = 0.207$; $F(1,97) = 1.263$, $p = 0.121$]. For the wildlife community dataset, forward selection resulted in two optimal RDA models, which accounted for similar proportions of variance ($R^2_{adj} = 0.201; 0.197$). The first model (W1) included the terms habitat diversity and proportions of grassland and tree cover, and the second (W2) included the terms habitat diversity and proportions of bareground, grassland, shrubs and artificial land use (all terms except bareground were statistically significant) (Table 4.3). In W1, the presence of primates, carnivores and fruit bats was

strongly correlated with tree cover in households, and in W2 was negatively correlated with proportions of artificial land use in households (Figure 4.4). In respective models, presence/absence of all avian functional groups except seed-eating and scavenging birds were correlated with households containing a larger proportion of trees, grassland or shrubs, or households with more diverse habitats, and negatively correlated with artificial land use. Communities dominated by rodents, scavenging and low-strata seed-eating birds and insectivorous bats were negatively correlated with environmental variables (trees, grassland, shrubs and habitat diversity), and positively correlated with artificial land use. When coloured according to their SOM groupings, household modes showed clear differentiation in wildlife community structure and the environmental conditions determining this structure (Figure 4.4). Environmental niches explained all variation in the community assemblage in the model W1, whilst anthropogenic and environmental niches explained similar variance in community structure in model W2 (W1: 0|0.201, W2: 0.149|0.129) (Table 4.3). Environmental and anthropogenic factors were associated with variation in avian community structure across households.

Optimal RDA models accounted for similar proportions of variation in avian community structure (R2adj = 0.267, 0.23). Both models contained habitat diversity and abundance of pigs, whilst model A1 also included proportions of grassland and trees, and model A2 included elevation and proportion of artificial land use (all terms except elevation were statistically significant) (Table 4.3). Households with lower proportions of shrubs and tree cover, more homogenous biotic habitats and higher proportions of artificial land use were more likely to have species assemblages dominated by scavenging and low-strata seed-eating birds (generalists), whilst households with more trees, diverse biotic habitats and shrubs were positively correlated with omnivorous, invertebrate-eating, and fruit/nectar birds occupying high and low strata (specialists) (Suppl. Figure 4.4). Low-strata invertebrate-eating birds were associated with households with a higher proportion of grassland and shrubs, greater habitat diversity, and higher abundances of pigs, whilst in model A2, generalist invertebrate-eating birds were also associated with pig abundance. Once split into fractions, environmental variables accounted for significantly larger proportions of variance in avian community structure in both models (A1: 0.02|0.267, A2: 0.126|0.159) (Suppl. Table 4.2).

	Optimal RDA					Optimal dbRDA		Optimal Linear Mixed Effects Models	
	Avian [1]	Avian [2]	Wildlife [1]	Wildlife [2]	Livestock	WL [1]	WL [2]	LCBD [1]	LCBD [2]
R^2	*0.267*	*0.23*	*0.201*	*0.197*	*0.041*	*0.092*	*0.124*	*0.253*	*0.253*
Environmental variables									
Habitat Diversity	**0.004**	**0.014**	**0.004**	**0.014**	NA	NA	NA	**-0.0056; 0.0007**	**-0.0056; 0.0007**
Elevation	NA	0.901	NA	NA	NA	NA	NA	NA	NA
% Grassland	**0.002**	NA	**0.014**	**0.032**	NA	NA	NA	NA	NA
% Shrubs	NA	NA	NA	**0.001**	NA	NA	NA	NA	NA
% Trees	**0.001**	NA	**0.001**	NA	NA	0.135	NA	NA	NA
% Waterbody	NA	NA	NA	NA	NA	NA	NA	NA	NA
Anthropogenic variables									
% Artificial	NA	**0.014**	NA	**0.024**	NA	NA	**0.015**	NA	NA
% Bareground	NA	NA	NA	0.093	NA	NA	NA	NA	NA
% Cropland	NA	NA	NA	NA	NA	NA	NA	NA	NA
Garbage	NA	NA	NA	NA	NA	NA	NA	NA	NA
Manure	NA	NA	NA	NA	NA	NA	NA	NA	NA
Total cattle*	NA	NA	NA	NA	NA	NA	NA	NA	NA
Total goats*	NA	NA	NA	NA	NA	NA	NA	NA	NA
Total indigenous chickens*	NA	NA	NA	NA	NA	NA	NA	NA	NA
Total lagomorphs	NA	NA	NA	NA	NA	NA	NA	NA	NA
Total pigs*	**0.007**	**0.005**	NA	NA	NA	NA	NA	NA	NA
Total poultry (other)*	NA	NA	NA	NA	NA	NA	NA	NA	NA
Total sheep*	NA	NA	NA	NA	NA	NA	NA	NA	NA
Total waterfowl*	NA	NA	NA	NA	NA	NA	NA	NA	NA
Ruralness Index	NA	NA	NA	NA	NA	NA	NA	NA	NA
Wealth Index	NA	NA	NA	NA	**0.012**	**0.01**	**0.005**	-0.0002; 0.1215	-0.0002; 0.1215
Household Area	NA	NA	NA	NA	**0.016**	NA	NA	NA	NA
dbMEM	NA	NA	NA	NA	Cond	Cond	Cond	-0.0007; 0.0564	-0.0007; 0.0564

Table 4.3: R^2_{adj} values for global and optimal redundancy analysis (RDA), distance-based redundancy analysis (dbRDA), and linear mixed effects models (LMM), examining household determinants of community structure for different host assemblages. P-values for variables included in these models are also shown (values in bold signify statistically significant fixed effects). Coefficients and p-values are depicted for linear mixed effects models. Wildlife-livestock is abbreviated to WL; [1] denotes model including tree cover (and not artificial land use); [2] denotes model including artificial land use (and not tree cover); NA (black) means that the variable was not included in the model; NA (red) means that the variable was included in the global model only, and not after selection for the optimal model; * indicates that variable was square root transformed before inclusion in the model.

Livestock. Variation in household livestock keeping assemblages was associated with wealth and household area (R2adj = 0.04) (Table 4.3).

Wildlife and Livestock. Variation in β–diversity of household wildlife-livestock communities was best described by artificial land use and household wealth index, which explained 12.4% of variance between household community (dis)similarity (R2adj = 0.124). As such, measurable differences in host structure between households

is either determined by variation in household wealth, or proportion of artificial land use (Figure 4.5). Households belonging to the same SOM mode grouped together, supporting earlier results that suggest modal groups are defined by similar wildlife-livestock community structures. Households belonging to modes 1 and 2 (high density, low α–diversity) were associated with high proportions of artificial land use and lower wealth indices, whilst those belonging to modes 3 and 4 (low density, high α–diversity) were associated with lower proportions of artificial land use and higher wealth indices.

Compositional stability of household interfaces. Household LCBD values (measuring the degree of 'uniqueness' of the host community within each household) were associated with changes in habitat diversity, wealth index and the spatial eigenvector MEM10 (which represents spatial variation across medium spatial scales) (marginal R2: 0.253, Table 4.3). Habitat diversity was the only statistically significant term in this model, and was negatively correlated with LCBD (β = -0.006, 95% CI = -0.009 – -0.002, $P < 0.001$) (Figure 4.5). As the diversity of ecological habitats within households decreases (and thus the host biotic niche becomes less complex), LCBD increases, and host assemblages become more unique.

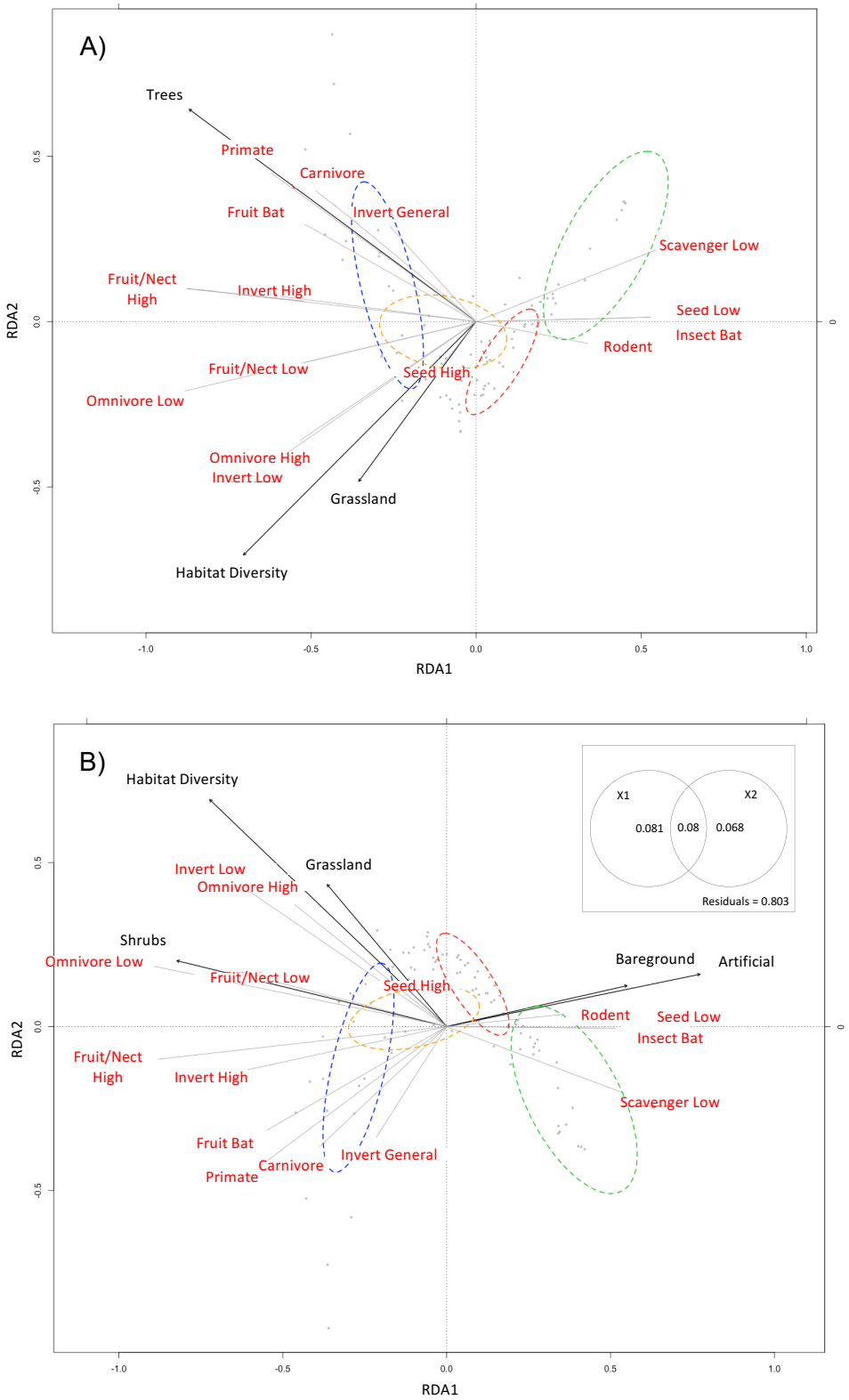


Figure 4.4: Correlation triplots for optimal redundancy analysis (RDA) models examining determinants of community structure for wildlife assemblages (models W^1 and W^2. **A)** Model including tree cover rather than artificial land use (response variables red, explanatory variables black), **B)** model including artificial land use rather than tree cover (response variables red, explanatory variables black). Grey dots represent households. Ellipses represent household modal membership (as determined in the broad scale self-organising map (SOM)); green = 1, red = 2, orange = 3, blue = 4. Box in the upper right corner of plot **B)** indicates R^2_{adj} values for each set of explanatory variables (derived from variance partitioning), where X1 = anthropogenic determinants, and X2 = environmental determinants.

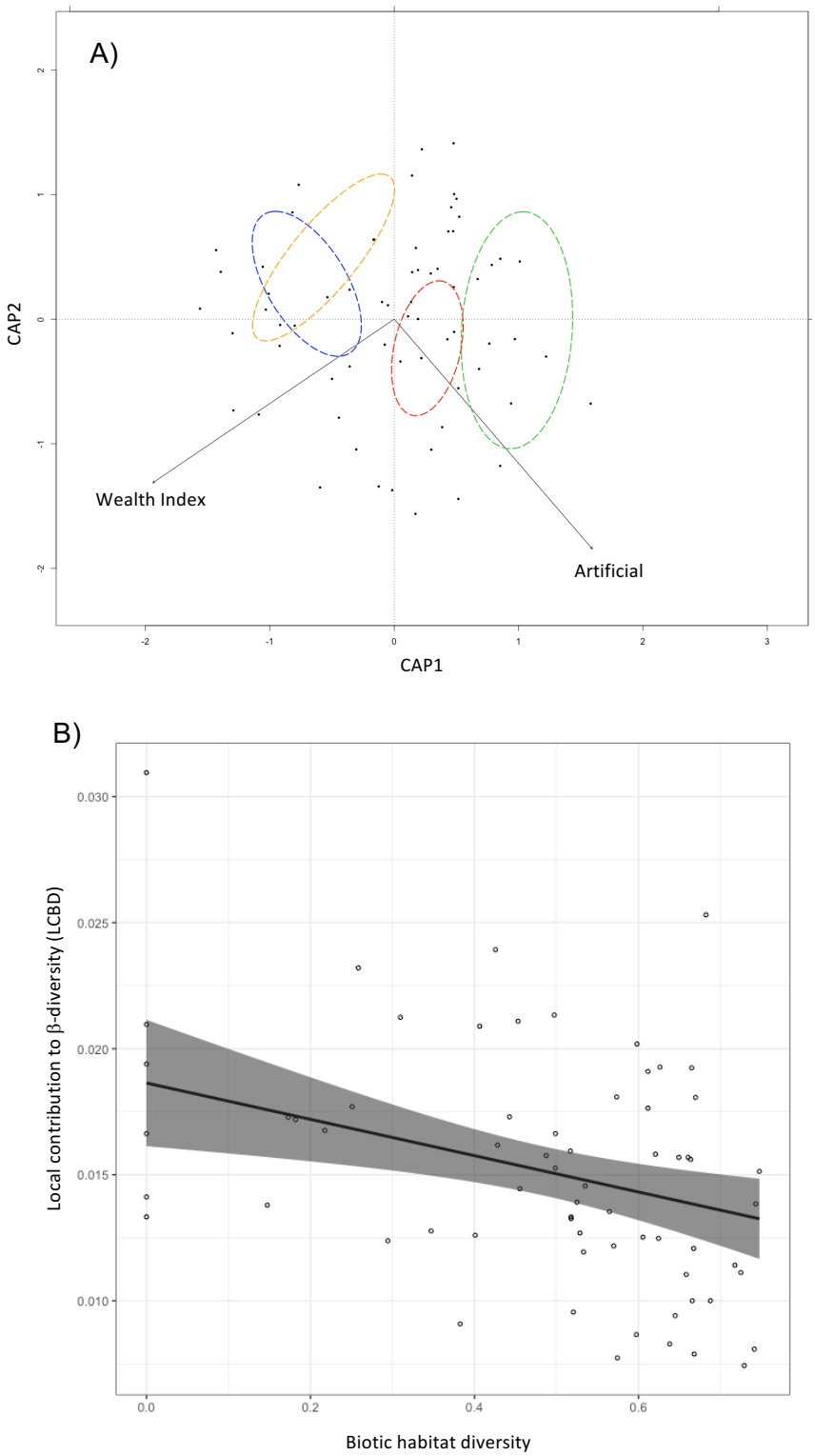


Figure 4.5: Plots for distance-based β-diversity analysis **A)** Correlation triplot of optimal distance-based redundancy analysis (db-RDA) model examining determinants of dis-similarity in household wildlife and livestock communities (WL^2). Response variables [principle coordinates representing the (dis)similarities between household assemblages] are represented as black points, explanatory variables are black lines and text. **B)** Fit of the linear mixed effects model examining determinants of household LCBD values, demonstrating the effect of household biotic habitat diversity on LCBD value. LCBD values provide a measure of how unique households are in terms of their wildlife-livestock assemblages. All other covariates in the model are kept constant.

4.4.3 Zoonotic parasite richness and sharing in wildlife with synanthropic traits and in livestock

Rodents had the highest estimated zoonotic parasite richness of all wildlife groups (262 species of parasites), followed by primates (124), carnivores (122), fruit bats (73), invertebrate-eating birds (64), insectivorous bats (63), seed-eating birds (53), omnivorous birds (25), scavenging birds (14) and fruit/nectar eating birds (6). Livestock hosted an estimated 508 zoonotic parasites (Suppl. Table 4.2). Rodents also shared the highest number of zoonotic parasites with other taxa (estimated degree centrality: 456), and were thus most central when functional groups were plotted as a network of zoonotic parasite sharing (Figure 4.6). Livestock and other mammalian groups (fruit/nectar and insectivorous bats, carnivores and primates) were also relatively central in the network (estimated degree centrality: 395, 311, 306, 297, 295 respectively), whilst seed-eating, omnivorous and insectivorous birds were the most central avian groups in the network (estimated degree centrality: 233, 193, 192 respectively) (Suppl. Table 4.2). Due to large variation in the number of species sampled and research effort per functional group (fruit/nectar eating birds were particularly under-represented), these results are relative and should be interpreted with caution.

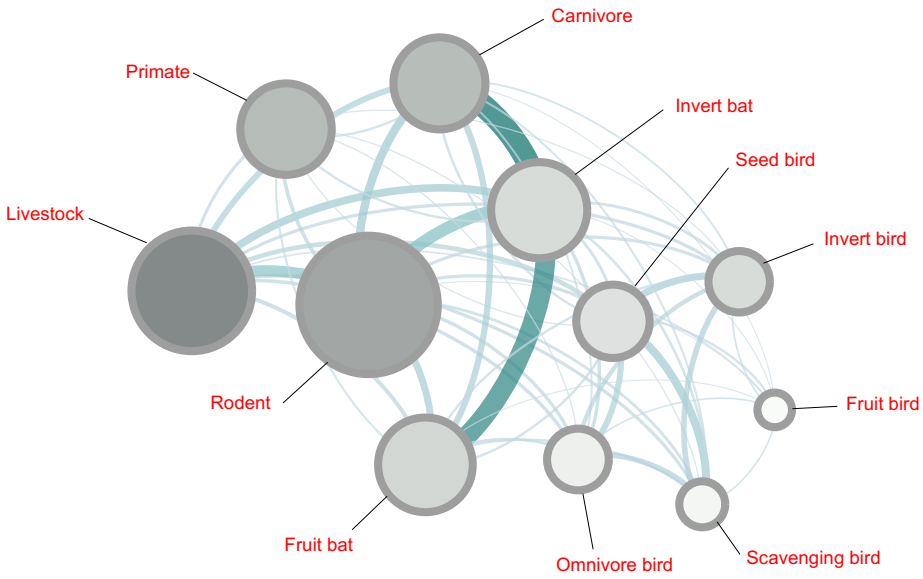


Figure 4.6: Network of zoonotic parasite sharing between wildlife functional groups showing synanthropic traits, and livestock. Nodes (circles) represent functional groups, and are shaded according to asymptotic estimates of zoonotic parasite richness (light – dark). Size of nodes is comparable to asymptotic estimates of host sharing (degree centrality, small – large). Edges linking nodes are weighted according to the raw number of zoonotic parasites shared between nodes (thin-thick).

Raw data pertaining to the number of zoonotic parasites shared between individual species was also plotted as a network, revealing finer scale details of zoonotic parasite sharing associations between functional groups (Figure 4.7). On examination of this network, host tropism of parasites was evident, with parasite sharing clearly structured by functional groups of hosts, nested within mammalian and avian classes. However, a single group, comprising pigs and a limited number avian and mammalian species representative of all wildlife functional groups except fruit/nectar eating birds, demonstrated a higher level of competency as hosts for both avian and mammalian zoonotic parasites.

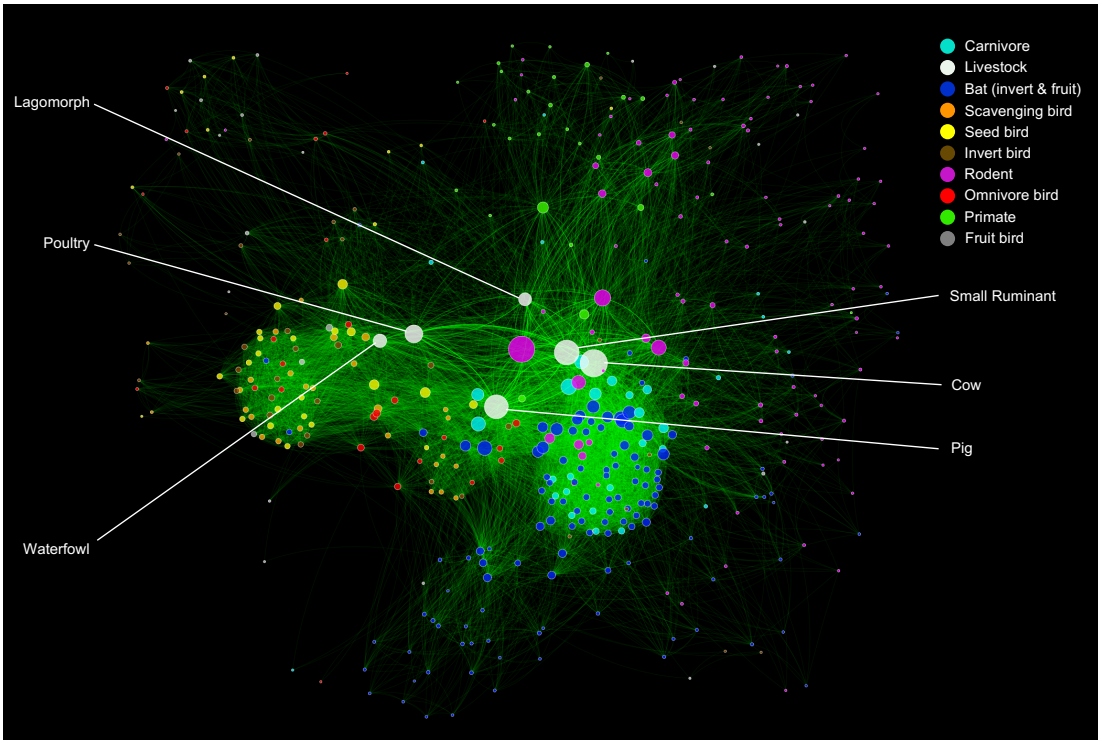


Figure 4.7: Network of zoonotic parasite sharing between wildlife species that utilise human environments and livestock taxonomic groups. Nodes are coloured according to membership of wildlife functional groups and livestock. Each edge is unweighted, and represents a single host-host parasite association. Size of each node represents number of parasites hosted by each taxon.

4.5 Discussion

Species assemblages, as studied by ecologists, represent the best available measure to estimate the impact of anthropogenic change on ecosystems [110]. As such, studying the distributional ecology of host communities across gradients of land-use change is an essential component of understanding the influence that such ecological change has on pathogens. In this study, methods from community ecology were applied to describe how urban land-use change influences host community structure within households, which represent important wildlife-livestock-human interfaces. Following a scaled approach, the results reveal variation across multiple levels of urban biological organisation, and shed light on the impact of environmental and human factors on the structure of urban host communities.

Nairobi, like other urban environments, is characterized by high heterogeneity of land use, resulting from fragmentation of natural habitats and anthropogenic activities. Analyses show that habitat structure in households runs along a gradient of decreasing biotic complexity, with most variation being attributed to decreasing proportions of tree cover, shrubs and grassland, and increasing proportions of artificial land use. This change in land use from largely environmental (biotic) to predominantly anthropogenic (abiotic) habitats is to be expected, and has a profound impact on the community assemblage of wildlife species in urban environments [23]. As biotic habitat diversity decreases, and the proportion of anthropogenic, abiotic land use increases, the α-diversity of avian species and avian and mammalian functional groups declines. The wildlife groups that remain (and in the case of avian functional groups, increase in relative abundance) represent urban generalists, capable of utilising resources in a broad variety of environmental and anthropogenic niches. These synanthropic mammals and birds (rodents, scavenging and seed-eating birds, and insectivorous bats) are found ubiquitously in households across the city, and as such frequently co-occur with commonly kept livestock species (indigenous chickens and small ruminants), generating 'baseline' wildlife-livestock interfaces across the urban landscape. As habitat diversity and proportions of biotic land use in households increase, wildlife functional groups that utilise restricted niches (such as frugivores, nectarivores and primates) can survive, and host communities become more complex. Variance partitioning was used to compare the relative influence of environmental and anthropogenic niches on structuring wildlife community assemblages in households.

Without artificial land use in the models, environmental niches accounted for at least 90% of explained variation in avian and wildlife community structures, which dropped to 46% when artificial land use was included instead of tree cover. As such, the two sets of niches appear relatively evenly balanced in their effect on wildlife-host community structure in households across the city.

Evidence to suggest that resource-provision through anthropogenic activities within households was also responsible for structuring wildlife communities was not found. This is not to say that resource provisioning from livestock, their products or their feed does not affect the distributional ecology of wildlife; a significant body of evidence exists suggesting that this is not the case (reviewed in [86]), and the relationship between the abundance of species belonging to insectivorous avian functional groups and pigs in households in this study, could be due to variation in arthropod numbers associated with pig-keeping. However, without measuring wildlife species abundances (for both mammals and birds), it is possible that this dataset lacked the resolution to detect these patterns more widely. Collecting such data was beyond the scope of this study, but would have increased the resolution and discriminatory power of my analyses, by allowing computation of α–diversity statistics that account for both richness and evenness of the wildlife community, and enabling me to conduct multivariate analysis on a community dataset consisting of abundance data, which is more information-rich, and less biased in the face of under sampling [153].

The results of this study demonstrate that the structure of wildlife and livestock communities in Nairobi depend on very different sets of determinants. Unlike wildlife, livestock are uncoupled from resource-based dependency on a niche, with their distribution and community structure being determined by a complex set of social anthropological factors; wealth and household area had significant effects on the relative abundance of livestock species in households. As such, the true determinants of wildlife-livestock community structure across urban households lie in a combination of environmental/anthropogenic niches and social factors. Evidence to support this assertion was apparent in analysis of β-diversity in wildlife-livestock community structure between livestock keeping households. Differentiation in host community structure was determined by wealth (as a driver of livestock community structure) and artificial land use (representing the continuum of habitat niche structure); households of similar wealth or with similar proportions of artificial land use typically harboured similar assemblages of hosts, and thus represented interfaces

of broadly congruent structure. The uniqueness of household host assemblages (LCBD indices) was negatively associated with biotic habitat diversity, indicating that host community structures were more unusual (or 'unstable') at the anthropogenic end of the urban land-use spectrum. This result is at odds with other studies, that have identified a trend towards 'biotic homogenisation' (wildlife communities becoming more structurally similar) under increasing levels of urbanisation [154]. It is possible that this observation is explained by the extreme variation in types of urban development in a developing city such as Nairobi, generating variability in abiotic niches and thus in the communities that utilise them. Using self-organised maps (SOMs), these findings were related to the household scale, and used to generate meaningful classifications of household interfaces across the urban landscape. The results are four groups of household interface that represent community organisation across multiple scales (Table 4.2). Households belonging to mode 1 lie at the 'anthropogenic extent' of urban land use, representing interfaces with low biotic habitat diversity that are dominated by abiotic anthropogenic niches. These interfaces support wildlife communities of low species and functional diversity dominated by urban synanthropes. Here, wildlife co-exist with a high density of humans, and high density, low or medium diversity livestock communities, that are characterised by poultry, small ruminants and rabbits. At the opposite end of the scale are households representing more ecologically 'normal' conditions, featuring diverse biotic habitats, with environmental niches populated by progressively more functionally complex wildlife communities. Wildlife in these communities co-exist with low densities of humans and livestock.

Biodiversity loss has been linked to impairment of ecosystem function and changes in the epidemiology of disease, and as such the urban restructuring of host assemblages observed in this study has important implications for human health and wellbeing in Nairobi [155,156]. At any given wildlife-livestock-human interface, the biotic niche inhabited by a community of parasites is determined by the in-situ assemblage of their hosts, and as such, any perturbations in host community structure could have important effects on the ecology and epidemiology of these parasites. Taking two broadly accepted principals regarding the dynamics of parasite communities in relation to changes in their host niche, and relating these to the variation in host community structure in urban households described in this chapter, broad ecological and epidemiological trends in parasite dynamics across the urban

landscape can be inferred, and used to generate a set of testable hypotheses to be explored in future studies (Figure 4.8). These hypotheses can also be extended to the genomic scale, where the relationship between microorganisms and mobile genetic elements (mobile sequences of DNA, encoding virulence or antimicrobial resistance genes, that can move within or between microorganism genomes) can be thought of analogously to that of host and parasite. In generating these hypotheses the following assumptions were made: *i)* host and pathogen α-diversity are correlated (the exact nature of this relationship would depend upon host specificity – saturation is expected to occur more quickly when communities are dominated by parasites with low host specificity [157]); *ii)* changes in host relative abundance and density influence the structural stability of the parasite community.

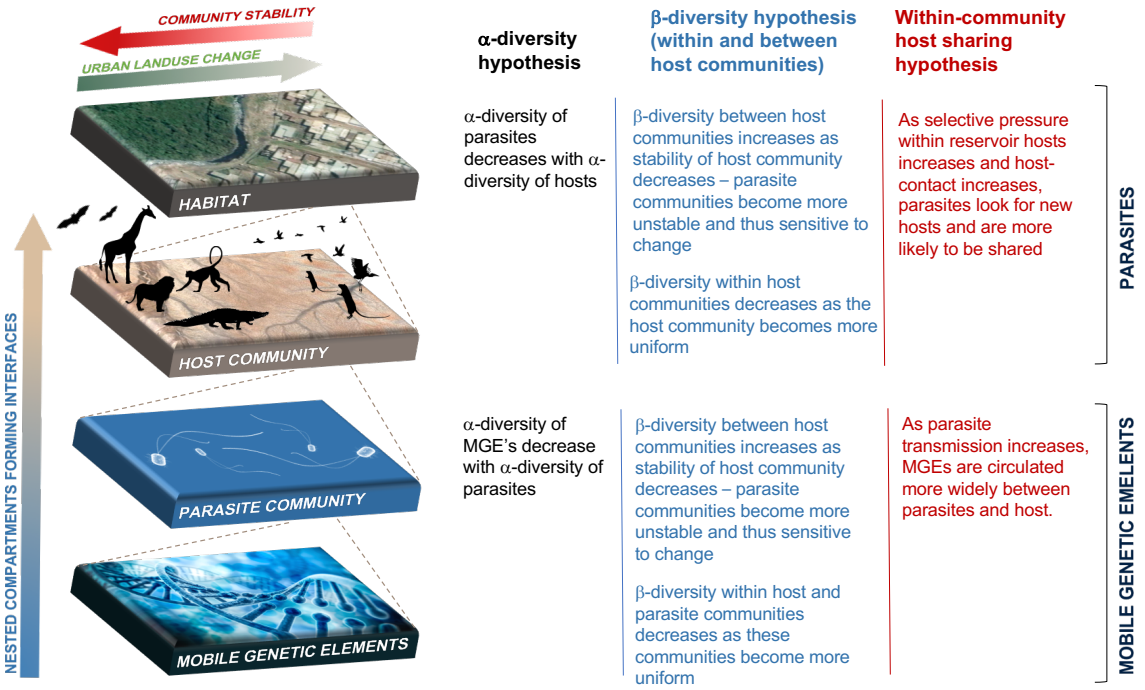


Figure 4.8. Schematic, demonstrating hypotheses relating urban land-use change to the dynamics of communities of parasites and the mobile elements of their genomes.

As the most obvious trend in host community structure across households, It is reasonable to assume that as wildlife communities become more functionally uniform (and livestock and human density increases), major structural changes will occur in their microbial communities. At the biotic end of the urban land-use spectrum, microbial parasites exist within a diverse, relatively stable community of hosts, and as such, would be expected to belong to equally diverse and stable communities (resulting in high microbial α-diversity, high β-diversity between microbial communities constrained within interfaces, and low β-diversity for microbial communities between interfaces). At the, 'anthropogenic', end of the urban spectrum, microbial communities existing within a restricted host niche would be expected to face higher selection pressures, and thus be less structurally stable (with conditions favouring the persistence and amplification of synanthrope-borne microbial parasites). Here, we hypothesise that communities are characterised by low β–diversity between microbial communities within interfaces (as α–diversity of hosts decreases and microbial communities become more homogenous), and higher β–diversity between interfaces (as the host structure of interfaces becomes less stable). Given that interfaces at the 'anthropogenic' end of the spectrum tended to have higher densities of livestock and humans, such conditions could present ideal circumstances for microbial spillover from wildlife to occur.

Whether, and how, the risk posed to humans through exposure to a 'pool' of zoonotic pathogens varies with such urban epidemiological gradients is unclear. At a global scale, zoonotic viral richness is correlated with total viral richness within hosts, suggesting that if hosts existing in communities with higher host species α-diversity do indeed harbour greater microbial diversity, they should also host more zoonotic pathogens [135]. However, as demonstrated in this study among others [5,125,135], zoonotic parasite diversity is not equally distributed between host taxa, making the abundance of taxa that are competent hosts for a greater diversity of zoonoses (such as rodents, primates, carnivores and bats) important in determining the structure of the 'zoonotic pool' at urban interfaces. Host transmission potential (degree centrality: a measure of the host specificity of zoonotic parasites present with each host) would also be expected to influence the structure and dynamics of zoonotic parasite communities. In this study, wildlife functional groups that demonstrated synanthropic responses to urban land-use change in Nairobi also exhibited high transmission potential relative to

other functional groups and livestock. Rodents had the highest diversity and transmission potential of zoonotic parasites compared to other wildlife functional groups and livestock (supporting the notion that they act as important urban reservoirs for zoonoses [28,125]), insectivorous bats exhibited high transmission potential, and seed-eating birds had the highest transmission potential of avian functional groups. That these species respond to, and are present across, gradients of urban change, should make them a priority for future studies investigating the impact of urban change on zoonotic parasite carriage in wildlife. Systematic characterisation of the pathways linking host and microbial communities in a range of urban environments will be required to understand the epidemiological consequences of urban environmental change. Such studies are necessary to inform urban development planning that safeguards both human and animal health.

4.6 Conclusion

Households in Nairobi can be characterised by sympatric populations of wildlife, livestock and human hosts. Variation in wildlife host assemblages were best described by habitat alteration, whilst changes in livestock assemblages were described by wealth and household area. These patterns suggest that structural changes in urban host communities are intrinsically linked to abiotic drivers associated with urban land-use change.

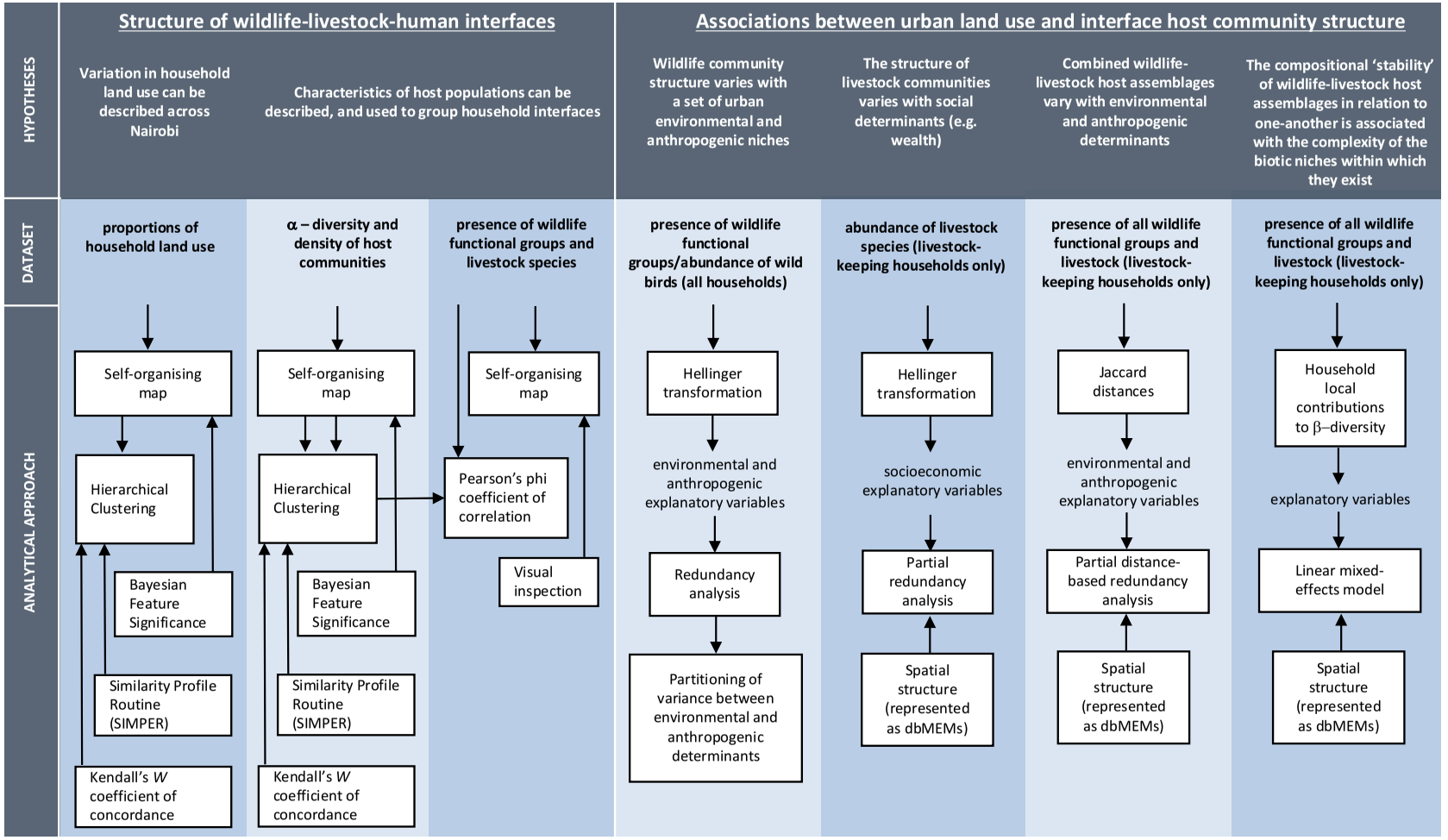


Figure 4.1: Flow diagram describing the analytical approaches used to test the hypotheses in this study. The first set of analyses focus on characterising the structure of wildlife-livestock-human interfaces in Nairobi, whilst the second set of analyses examine associations between variation in land use and host community structure at these interfaces.

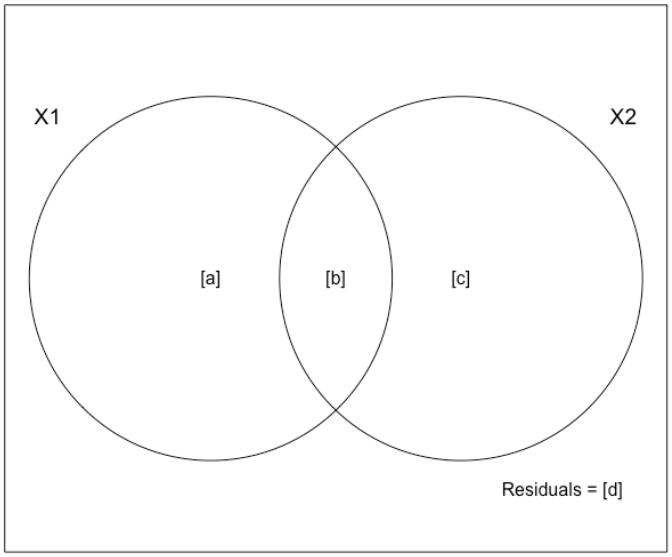


Figure 4.2: Variance partitioning fractions, when considering two sets of explanatory variables. Throughout this study, X1 relates to variation derived from anthropogenic determinants, and X2 relates to variation derived from environmental determinants. [a] relates to the portion explained by anthropogenic determinants alone, [c] relates to the portion explained by environmental determinants alone, and [b] relates to the portion explained by both anthropogenic and environmental portions.

		Variance Component						
Community		(x1)	(x2)	x1 + x2	x1|x2	x2|x1	b	d
Avian [1]	Significance (p)	**0.02**	**0.005**	**0.005**	**0.025**	**0.005**		
	Variance	0.024	0.247	0.271	0.024	0.247	0	0.733
Avian [2]	Significance (p)	**0.005**	**0.015**	**0.005**	**0.005**	**0.005**		
	Variance	0.126	0.159	0.226	0.066	0.1	0.059	0.775
Wildlife [2]	Significance (p)	**0.005**	**0.005**	**0.005**	**0.005**	**0.005**		
	Variance	0.149	0.129	0.197	0.068	0.081	0.081	0.803

Table 4.1: Variance partitioning results, performed on redundancy analysis (RDA) models. X1 relates to variation derived from anthropogenic determinants, and X2 relates to variation derived from environmental determinants. [1] denotes models with tree cover, and [2] denotes models with artificial land use.

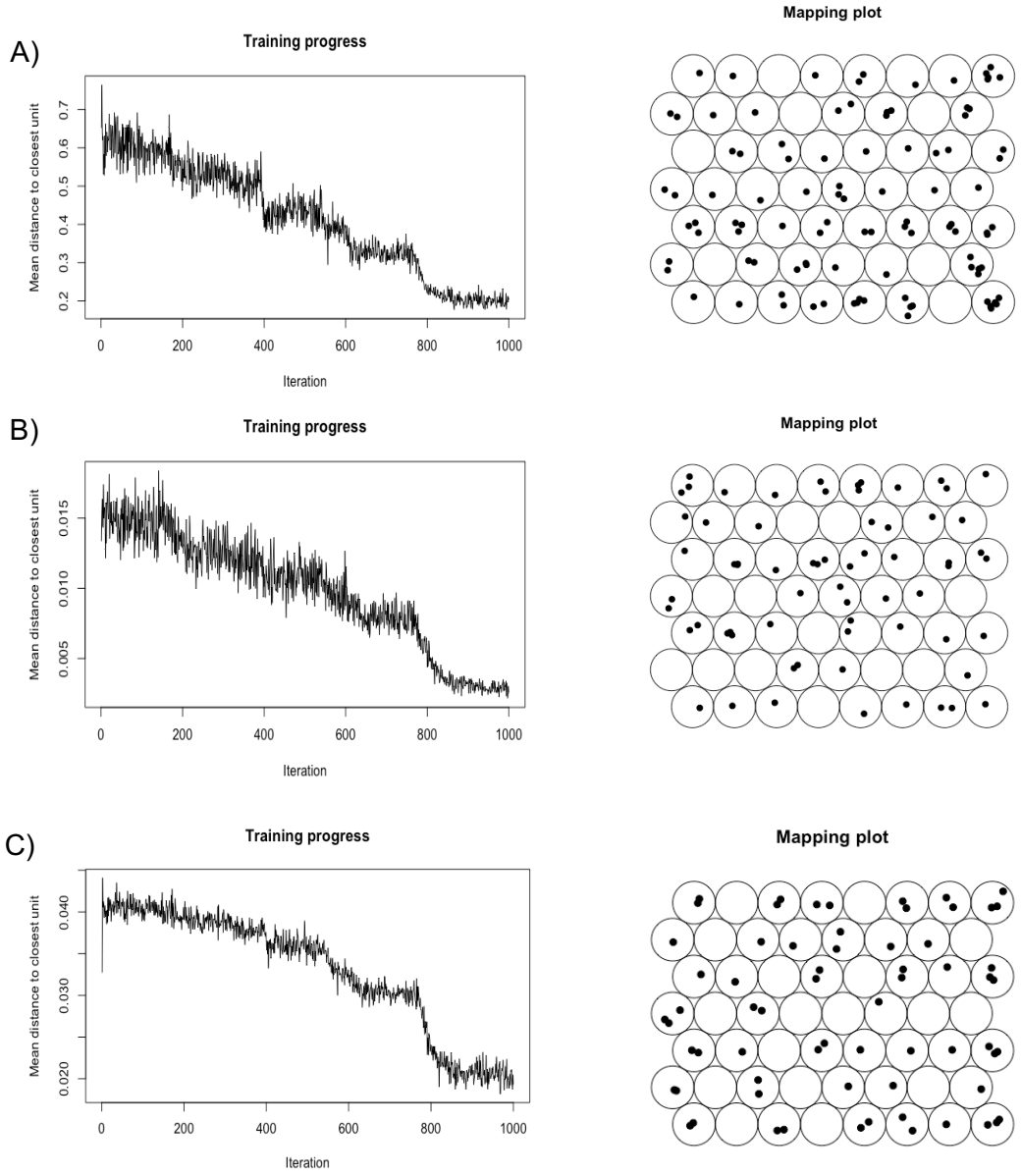


Figure 4.3: Diagnostic plots for Self-organising maps (SOMs). **A)** Training and mapping plots for land use SOM, **B)** training and mapping plots for SOM mapping broad patterns of household community structure, **C)** training and mapping plots for SOM mapping fine scale patterns of household community structure.

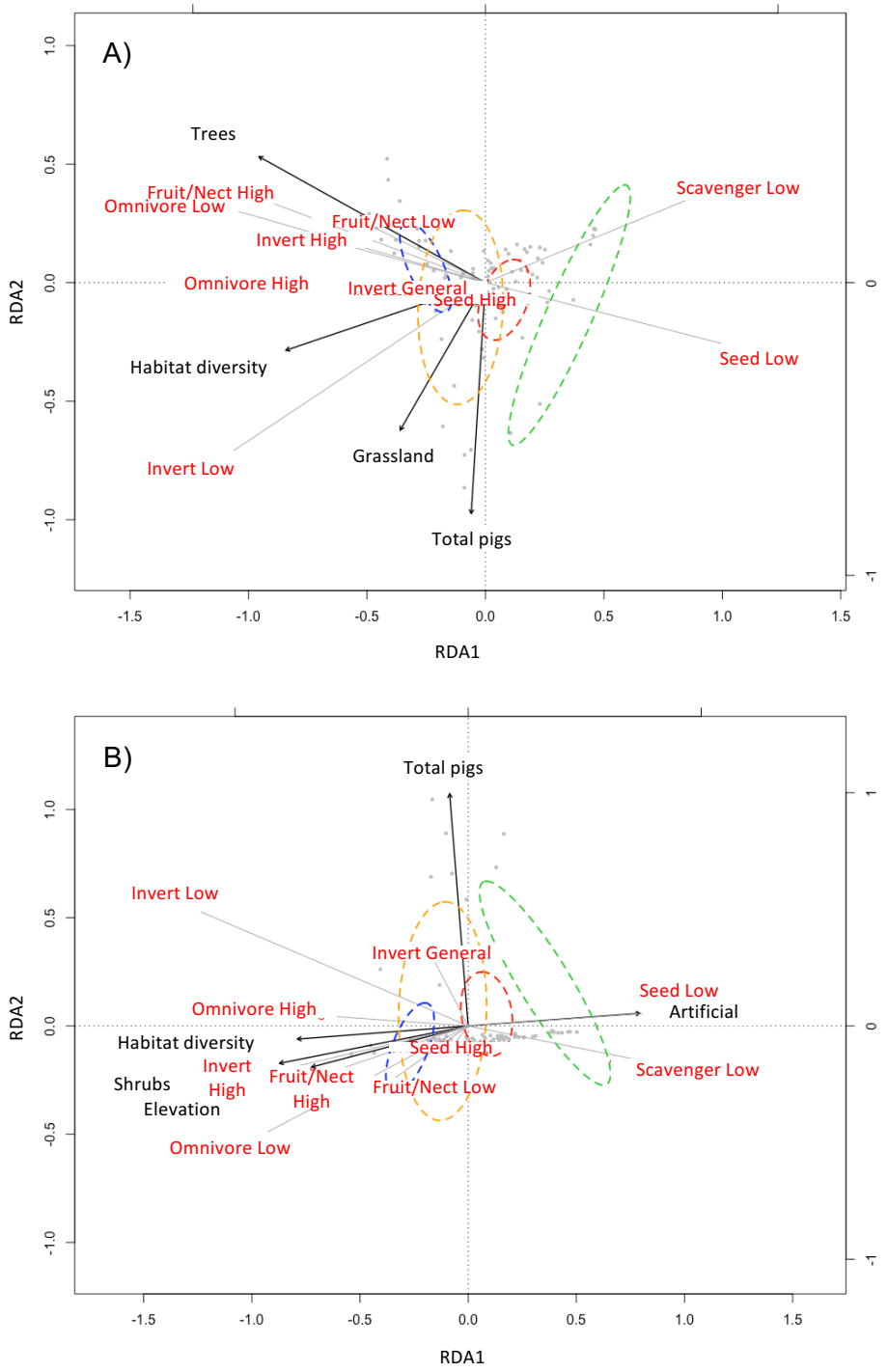


Figure 4.4: Correlation triplots for optimal redundancy analysis (RDA) models of the avian community dataset (models A^1 and A^2). **A)** Model including tree cover rather than artificial land use (response variables red, explanatory variables black), **B)** model including artificial land use rather than tree cover (response variables red, explanatory variables black). Grey dots represent households. Ellipses represent household modes (as determined in the broad scale SOM); green = 1, red = 2, orange = 3, blue = 4.

Functional Group	Asymptotic parasite diversity	Asymptotic degree centrality	Species (n)	Hindex
Livestock	508	395	11	89
Rodents	262	456	107	39
Insect bats	63	311	89	16
Primates	124	295	28	5
Carnivores	122	297	33	8
Fruit bats	73	306	38	8
Seed-eating birds	53	233	34	12
Fruit/Nectar bird	6	100	7	2
Omnivorous birds	25	193	28	5
Invertebrate-eating birds	64	192	38	7
Scavenging birds	14	139	27	7

Table 4.2: Statistics derived from analysis of the host-parasite association database. Asymptotic estimates for parasite diversity and degree centrality were derived using Chao2 non-parametric models. Species (n) shows the number of species belonging to this functional group represented in the database. Hindex represents a measure of sampling effort for each functional group.

Chapter 5

Urban Wildlife and the Epidemiology of Antimicrobial Resistance in Nairobi

5 Urban Wildlife and the Epidemiology of Antimicrobial Resistance in Nairobi

5.1 Abstract

Antimicrobial resistance (AMR) represents one of the great challenges facing global health security in the modern era. Wildlife species, particularly those that utilise urban environments, are an important but understudied component of AMR epidemiology. In this chapter I use a large bacterial dataset to examine the role of wildlife in the spread of clinically relevant resistance across the developing city of Nairobi. Faecal samples (n=2102) were collected from 78 wildlife species (n = 849), 13 livestock species (n=656), humans (n=333) and the external environment (n=288) from 99 households across Nairobi. *E. coli* was cultured, and a single isolate from each sample tested for sensitivity to 13 antibiotics. Statistical models were developed to answer three questions; *i)* when compared to humans, livestock and the environment, are urban wildlife a net source for antimicrobial resistance in Nairobi? *ii)* what is the prevalence of AMR phenotypes and multi-drug resistant (MDR) *E. coli* carriage in urban wildlife, and is this linked to variation in ecological traits, such as foraging behaviour? *iii)* what are the household-level risk factors for sharing of AMR between humans, wildlife and livestock? *E. coli* were isolated from 485 wildlife samples. Wildlife carried a high prevalence of clinically relevant AMR-*E. coli* [90.7% (440/485) resistant to at least one antibiotic] and multi-drug resistant (MDR) [52% (252/485)], which varied between taxa and by foraging traits. Extensively-drug resistant (XDR) and pan-drug resistant (PDR) isolates were also present. The phenotypic diversity of AMR-*E. coli* in wildlife was lower than in livestock, humans and the environment. Within household compounds, statistical models identified two "interfaces" for AMR exchange; *i)* between rodents/seed-eating birds, humans and their rubbish, and *ii)* between rodents/seed-eating birds, cattle and bovine manure. Our results are consistent with the hypothesis that wildlife are a net 'sink' rather than source of clinically relevant resistance. These results provide novel insight into the determinants of AMR carriage in wildlife, and routes of exposure between wildlife and humans, livestock and their shared environment.

5.2 Introduction

Antimicrobial resistance (AMR) represents one of the great challenges facing global health security in the modern era, and will ultimately limit our capacity to treat microbial infections. The repercussions for human and domestic animal health are severe; as infections become more difficult and costly to treat, morbidity and mortality will increase, and the extra burden placed on health services and livestock production will have considerable economic consequences [158].

Misuse of antibiotics is common, particularly in the agricultural sector where productivity-driven intensification (in both terrestrial and aquaculture systems) relies heavily on the use of antibiotics to increase production and manage animal health and welfare [159,160]. Unsurprisingly, the two most likely sources of clinically relevant AMR and resistance genes are pathogens exposed to antibiotic usage in humans and livestock [161], but it is increasingly clear that focussing on these compartments of the transmission system, at the expense of AMR dispersal in the wider environment (such as wild animals, water and soil), will result in an incomplete epidemiological picture of resistance [162]. Due to its ubiquity in bacteria, AMR represents a multiple-agent multiple-host system, and as such, more than any single pathogen, the study of resistance epidemiology is suited to an intersectoral ecosystem health approach. As clinically relevant AMR from domestic animals and humans is released into the environment, landscape-level changes to biomes and their habitats could have knock-on effects for how resistant bacteria and genes are dispersed in the environment [121]. As Robinson et al. [160] speculate, lax environmental legislation and unregulated antibiotic usage might render these factors more pronounced in developing countries. The epidemiology of AMR in the city of Nairobi, which represents a developing country urban system, is now considered, simplifying the complexity of such top-down systems into a series of interlinked compartments.

5.2.1 Resistance Compartments

Humans and Livestock

Humans and livestock are regarded as the primary source of most clinically relevant AMR bacteria and resistance genes due to selective pressures on bacterial

populations following antibiotic use, although there is limited evidence with which to qualify the threat to human health posed by antimicrobial use in livestock and vice-versa. Poor-quality antibiotic drugs and substandard regulation, aberrant prescribing and dispensing practices, and self-medication can all promote selection for resistance genes in humans and livestock in developing countries [163,164]. Similar processes are likely to take place in clinical residues such as agricultural waste, sewage and hospital waste, where high concentrations of resistance determinants are found. In the human gut, direct exposure to resistance determinants can occur through the consumption of animal-sourced food products, and contaminated vegetation [165].

Soil and water

Bacterial populations in aquatic and soil habitats are enormously diverse, and play critical roles in nitrogen cycling, carbon sequestration, and the stability of aquatic ecosystems [166]. These bacteria also act as reservoirs of naturally occurring bacterial resistance, the burden of which is exacerbated by flows of resistance elements and other chemicals (such as heavy metals) from livestock and human waste, that can co-select for drug resistance [167]. Resulting changes to microbial diversity could lead to damaging effects on terrestrial and aquatic ecosystems, such as nutrification and mobilisation of heavy metals [168,169]. Rivers, and the movement of people, livestock, their products and wildlife, act as networks of connectivity in urban areas, and as such could represent important vectors for the dispersal of resistance elements within urban areas and to peripheral landscapes [9,170].

Wildlife

Wild animals could disperse AMR across urban environments and into peripheral landscapes. Ecological traits, such as habitat, feeding preferences and ranging behaviour determine the exposure of wildlife species to AMR, and how widely it is dispersed in the environment [171]. As land-use change reduces the availability of natural habitats to wildlife species, they are forced to seek alternative sources of food and shelter, bringing them into closer association with humans, livestock and their by-products, and increasing the potential for transfer of AMR between them [9,162]. For example, in urban areas, water birds have become increasingly dependent on irrigated farmland and wastewater treatment plants in the absence of natural wetlands [172]. Several studies have documented greater diversity and/or prevalence

of AMR in wildlife species that associate with human activities, which, when considered with the detection of extended-spectrum beta-lactamases (ESBL, enzymes that confer resistance to the highest priority, critically important antibiotics) is suggestive of direct spillover to wild animals from anthropogenic sources [173] (reviewed in [121,171]). The presence of diverse bacterial resistance profiles in wildlife inhabiting relatively pristine environments also demonstrates the complexity of naturally occurring AMR communities in the gut of free-ranging vertebrates, for which environmental acquisition probably plays an important role [121,174]. As such, it has been suggested that wild animals could represent a bioindicator, or sentinel for AMR in the environment [173,175]. The role played by urban wildlife as vectors for AMR dispersal is considered in Figure 5.1.

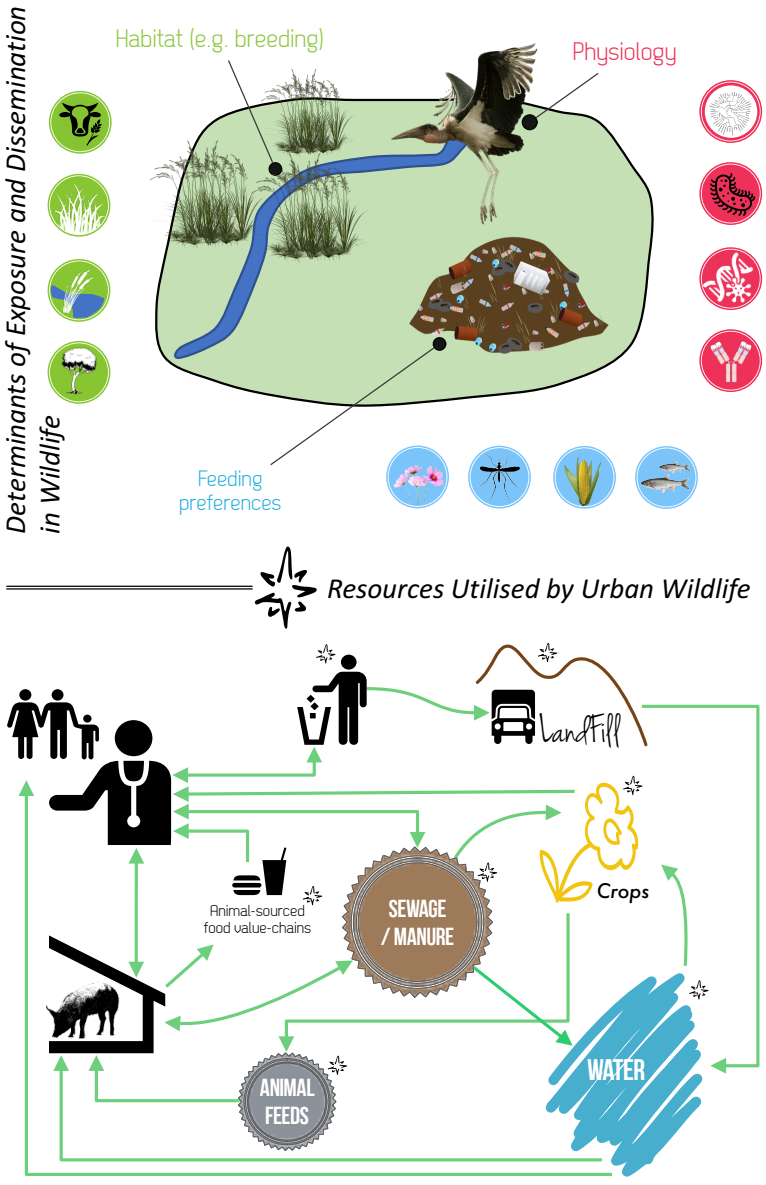


Figure 5.1. Graphic exploring the role played by urban wildlife in exposure to and dissemination of AMR determinants. The top panel ("Determinants of Exposure and Dissemination in Wildlife") describes three main groups of determinants that govern the carriage, exposure and dissemination of AMR in urban wildlife. The bottom panel ("Resources Utilised by Urban Wildlife", adapted from [121]) depicts the movement of AMR determinants in urban ecosystems. Stars denote compartments that urban wildlife are likely to utilise as resources, and thus potential routes of wildlife exposure to AMR.

The focus of this study is on informal livestock keeping in the developing city of Nairobi, Kenya, as a potentially high-risk urban interface for AMR dissemination into wildlife and the environment. In developing urban centres livestock are frequently kept within household compounds, where differing levels of waste management and anthropogenic resource provision could cause variation in environmental dispersal of AMR determinants, and exposure of wildlife to AMR [9]. As outlined above, wildlife are an important but often neglected component of AMR epidemiology, and as such attention in this study is focused on wild animals that utilise urban environments (synanthropes). Being ubiquitous in vertebrates and the environment, *Escherichia coli* is frequently targeted in studies of AMR, and represents an ideal bacterial candidate for studying AMR dispersal across diverse vertebrate host species and the environment [171].

Utilising *E. coli* AMR phenotypes collected from households across Nairobi, the role of urban wildlife in the epidemiology of antimicrobial resistance is explored. In considering only antibiotics with clinical use in human medicine, and clinically significant levels of resistance to these antibiotics (by using human treatment breakpoints), the clinical relevance of antimicrobial resistance in wildlife is examined. Wildlife, which are not treated with antibiotics, are expected to be a net recipient (or 'sink') of AMR in urban environments, and this hypothesis is tested by employing epidemiological and ecological models to compare the carriage of clinically relevant AMR between epidemiological compartments (wildlife, humans, livestock and the environment). To investigate determinants for carriage of AMR in urban wildlife, a scaled approach is taken. Recent studies have suggested the importance of resource utilisation in determining exposure of wildlife to AMR, and as such I expect host taxon and functional ecology (i.e. feeding traits) to play a role in wildlife AMR carriage. Variation in host taxon and functional ecology (i.e. feeding traits) are related to carriage of multidrug resistant (MDR)-*E. coli*, and AMR profile length in wildlife across the city. At a finer scale, epidemiological models are used to investigate risk factors for exchange of AMR between sympatric wildlife, livestock and humans, thus shedding light on pathways of epidemiological connectivity at household interfaces.

5.3 Methods

5.3.1 Samples and Laboratory Testing

Samples (n=2081) were collected from 78 wildlife species (birds and mammals, n=794), 13 livestock species (n=677), humans (n=333) and the external environment (n=277) in households across Nairobi, that were participating in the 99-household project (Chapter 3, Suppl. Table 3.1). An additional 24 samples were collected from birds and rodents in abattoirs across the city. Study design is explained in detail in Chapter 3.4 and [95], but briefly, Nairobi was split into administrative sublocations, and 33 were chosen on the basis of socioeconomic stratification. Three households were randomly selected to obtain two livestock keeping and one non-livestock keeping household per sublocation (a total of ninety-nine households), with the aim of maximising the spatial distribution and diversity of livestock keeping practices captured within the sampling frame. Wildlife, livestock and environmental samples were obtained from a single extra house, making the total number of households 100 in analyses for which human and questionnaire data were not considered. Households in each sublocation had to meet strict inclusion criteria of keeping small ruminants or poultry, large ruminants or pigs, or no livestock within the household compound. Abattoirs in Nairobi were selected and sampled as part of a separate value chain study conducted as part of the wider UrbanZoo project [176]. Wildlife samples were obtained by a range of taxon-specific trapping methods, which are described, along with protocols for collection of human, livestock and environmental samples in Chapter 3.5. Culture and isolation of *E. coli* isolates from samples is described in Chapter 3.6.

All isolates were revived and inoculated onto Mueller-Hinton plates prior to antimicrobial susceptibility testing. Isolates were then tested for susceptibility to ampicillin (10 μg), amoxicillin-clavulanic acid (30 μg), cefepime (30 μg), cefotaxime (30 μg), ceftazidime (30μg), chloramphenicol (30 μg), ciprofloxacin (5 μg), gentamicin (10 μg), nalidixic acid (30 μg), streptomycin (25 μg), sulfamethoxazole (30 μg), tetracycline (30 μg) and trimethoprim (2.5 μg) using the disc diffusion method according to the guidelines published by the Clinical and Laboratory Standards

Institute (CLSI [177]). CLSI guidelines were also used to determine the breakpoints for classifying isolates as 'sensitive', 'intermediate' or 'resistant' to the drug in question. Following previous studies, 'intermediate' strains were considered as resistant on an evolutionary basis, since these strains were deemed to be moving towards resistance [178,179]. Protocols for *E. coli* culture and ASTs were standardised across both laboratories. Multidrug-resistant *E. coli* (MDR) was defined as "non-susceptibility to at least one agent in three or more antimicrobial classes" following Magiorakos et al [178] (Suppl. Table 5.1). Wildlife isolates were also assessed for extensive drug resistance (XDR, "non-susceptibility to at least one agent in all but two or less antimicrobial classes") and pandrug resistance (PDR, "non-susceptibility to all antimicrobial agents tested") [178]. An antibiotic profile was defined as the combination of antibiotics to which an isolate is resistant, and thus profile length is the total number of antibiotics to which an isolate is phenotypically resistant.

5.3.2 Data analysis: Broad comparison between epidemiological compartments

Comparison of prevalence, multidrug resistance and profile length between compartments

A Bayesian analysis framework with MCMC was used to compare prevalence of resistance to 13 antibiotics between epidemiological compartments. This was applied using JAGS via the R package 'R2jags' [180,181]. A burn-in of 4000 iterations, three chains, and a thinning rate of 10 and 15,000 iterations for each posterior distribution was used. Diffuse normal priors $N(0,100^2)$ were used for the Bernoulli distribution, and the prior densities were plotted. Variation in prevalence between compartments was assessed by comparing posterior densities. Generalised Linear Mixed Effects Models (GLMMs) were developed to test whether MDR-*E. coli* carriage and AMR profile length differed between host compartments. The response variable for MDR-*E. coli* carriage was coded from 0 to 1, and profile length was measured from 1 to 13; as such, a binomial distribution with logistic link function and Poisson distribution were used, respectively. Spatial structure in the dataset was represented using distance-based Moran's eigenvector maps (dbMEMs), which

represent a powerful approach to model spatial structure in a response variable. dbMEMs are derived from spectral decomposition of the spatial relationships among sampling locations (for further details see Chapter 3.7.4) [110,111]. A set of dbMEM eigenfunctions modelling the spatial structure of the data at all scales were computed from the GPS coordinates of each sample using package 'adespatial' [112]. Linear trends were tested for and identified in the data, but spatial detrending was not applied before conducting dbMEM analysis (as recommended by Boccard et al [111]), because it was felt that these linear trends represented broadscale processes acting across the city, and were thus relevant to the models. Eigenvectors modelling positive spatial correlation were extracted, and regressed against each response variable in a generalised linear model (GLM). Step-wise selection was used to identify significant dbMEM eigenvectors from those modelling positive spatial correlation in the data. For interpretation, dbMEM variables (eigenvectors) were partitioned by broad, medium and fine spatial scales (Suppl. Figure 5.1). The data also consists of multiple observations within each household (or abattoir), which are nested within sublocations (see Chapter 3.7.3, Figure 3.3). As such, mixed effects models were applied with the random effects household, sublocation or both (the dependency structure of each model was assessed separately by fitting random-intercept models to the response data, and comparing the intra-class correlation coefficient [ICC, latent variable method [182]]). Sampling was carried out over the course of a year, and as such, temporal trends in the dataset were assessed by plotting mean MDR-*E. coli* carriage against the week in which samples were collected. Final models were constructed using stepwise, backwards elimination from the full model. Model assumptions were verified by plotting residuals versus fitted values, versus each covariate in the model. The residuals from each model were also assessed for spatial dependency, and plotted as a semivariogram in R package 'gstat' [183].

Ecological diversity

To assess how wildlife AMR profile diversity was distributed across all four epidemiological compartments (wildlife, livestock, human, environment), diversity was compared between compartments using four ecological measures of diversity related to Rényi's measures of generalized entropy. This approach is described by Mather et al. [1], where the exponential of Rényi's entropy ($D\alpha$) estimates the effective

number of species (AMR profiles in this case), and α represents a scale parameter, along which profile richness and relative abundance of profiles are weighted differently. The four diversity indices that were compared across compartments sit at different levels of α: at D0, profile richness (PR) is a count of AMR profiles (which ignores relative abundance, thus considering rare and common profiles equally); at log(D1), Shannon entropy (SEn) is the probability of any two isolates drawn at random having the same profile; at 1/D2 Simpson diversity (SD) is the relative abundance of each profile; and at 1/D∞ the Berger-Parker diversity (BP) is the proportion of the most common profile in the sample. Adjustments were made to each diversity measure to account for variation in sample size between compartments by resampling all compartments to the sample size of the smallest compartment (environment), with replacement. Bootstrapping 1000 times provided a median value with confidence intervals (CIs) for each of the diversity measures. For each index, the diversity index differences were compared between compartments using a Kruskal-Wallis test with Holm adjustment for multiple comparisons to compare ranges of α.

Profile richness and sampling estimates

Methods adapted from the community ecology literature were used to extend the comparison of phenotypic diversity between epidemiological compartments by estimating the number of undetected AMR profiles in each epidemiological compartment. This approach is similar to that of Anthony et al [184], who used comparable approaches to estimate mammalian viral diversity. Rarefaction and species accumulation curves were generated from the incidence data for all 13 antibiotics. The 'exact' accumulation method and non-linear regression models were used to estimate profile-area relationships, and thus the shape of the species accumulation curves. An appropriate non-linear model for each compartment was selected by subsampling half of the data, and extrapolating to the total number of samples in each compartment. Models with the closest predictions to the correct number of species were used. Sampling to the asymptote of these curves would reveal the total number of profiles in each compartment, but at prohibitive effort and costs. As such, Chao2, ICE and Jackknife incidence-based statistical methods were used to estimate the minimum total profile richness in each compartment from the data, by looking at frequencies of phenotype occurrence in collections of individuals. These

nonparametric estimators are considered more robust than model-based estimates, and particularly suited to the purposes of comparison between compartments, because they make no assumptions of distributions underlying species detection rates, which may differ between assemblages [185]. Rarefaction and species accumulation were performed in the R package 'vegan' [104], and richness estimates were computed in the R package 'fossil' [149]. To consider the implications for surveillance, methods from Chao et al [186] were followed to estimate the sampling effort required to detect a given proportion of the total AMR profiles estimated for each compartment.

5.3.3 Data analysis: role of wildlife functional ecology in AMR-*E.coli* carriage

A Bayesian analysis framework, as described in section 5.3.2, was used to estimate prevalence of resistance to 13 antibiotics between wildlife taxa. Ecological traits considered potentially important factors for exposure of wildlife to AMR were sourced from metadata and published sources (see Suppl. Table 5.2). These included taxonomic group (avian, bat, rodent), feeding ecology (food source and canopy strata) [97], home range, and association with water [187]. Home range estimates for all species except bats were calculated by allometric scaling of body weight [188]. Scaling factors published for functionally different mammals and birds by Ottoviani et al [189] were used, and species mean body weights were either collected during sampling, or sourced from published datasets when unavailable [190,191]. Two GLMMs were developed to investigate variation in the likelihood of multidrug resistant (MDR)-*E. coli* carriage and length of resistance profiles in wildlife, as determined by these functional traits. The response variable for MDR-*E. coli* carriage was coded from 0 to 1, and profile length was measured from 1 to 13; as such, a binomial distribution with logistic link function and Poisson distribution were used, respectively. For each set of response and explanatory variables, data exploration was carried out following the protocol described in Zuur et al [98] (for further details see Chapter 3.7.1). Data exploration revealed marked correlation between several of the explanatory variables, and as such, taxonomic groups and feeding niches were either combined or considered separately, and association with water was not considered as a covariate. Spatial structure was captured using db-MEMs, and the dependency structure of each model was assessed by fitting random-intercept models to the response data, as described in section 5.3.2. To account for the fact that antimicrobial sensitivity testing of wildlife

samples was split between two laboratories, the laboratory at which each sample was tested was included as a covariate in each model. The full models including all covariates are depicted in Suppl. Figure 5.2. Model selection procedures and checking of model assumptions were conducted as previously described (see section 5.3.2).

5.3.4 Data analysis: AMR exchange between wildlife, livestock and humans within households

A separate set of GLMMs were developed to investigate household-level risk factors for the likelihood of MDR-*E. coli* carriage in select urban wildlife with synanthropic traits (rodents and seed-eating birds). A set of anthropogenic and ecological covariates capturing AMR-*E. coli* carriage in livestock and humans, livestock keeping practices, land use within households and ranging behaviour of wildlife were derived from metadata and published sources (Suppl. Table 5.2). The proportion of cropland within each household was mapped by visual classification using ArcGIS, and subsequently ground truthed within each household (for further details see Chapter 3.5). Separate GLMMs with binomial distribution and logistic link function were fitted for rodents and seed-eating birds, using the same set of covariates. The response variable, MDR-*E. coli* carriage, was coded from 0 to 1. For each set of response and explanatory variables, data exploration was carried out following the protocol described in Zuur et al [98]. Spatial structure was captured using db-MEMs and the dependency structure of each model was assessed by fitting random-intercept models to the response data, as described in section 5.3.2. The full models including all covariates are depicted in Suppl. Figure 5.3. Model selection procedures and checking of model assumptions were conducted as previously described (see section 5.3.2).

5.4 Results

Samples were collected from a total of 547 individual birds, 9 'avian populations' (AP, 31 pooled samples), 167 rodents, 44 individual bats, 5 'bat populations' (20 pooled samples), 5 carnivores and 4 primates across 100 households and 11 abattoirs and butcheries in Nairobi. *E. coli* was isolated from, and AST's performed on, 51.6% (282/547) of birds, 64.5% (20/31) of AP's, 92.8% (155/167) of rodents, 50% (22/44) of bats, 22.2% (6/27) of bat populations, 60% (3/5) of carnivores and 100% (4/4) of primates. Due to low sample numbers, primates and carnivores were dropped from the dataset, and each pooled bat population sample was considered as coming from an individual bat for the purposes of all further analysis. *E. coli* was isolated from, and AST's performed on, a total of 638 livestock, 321 human and 256 environmental samples. In total, 52% (252/485) of samples from wildlife sampled in Nairobi carried MDR-*E. coli*, 1.6% (8/485) of wildlife isolates (all originating from birds) carried XDR-*E. coli*, and a single avian sample carried pandrug-resistant *E. coli*.

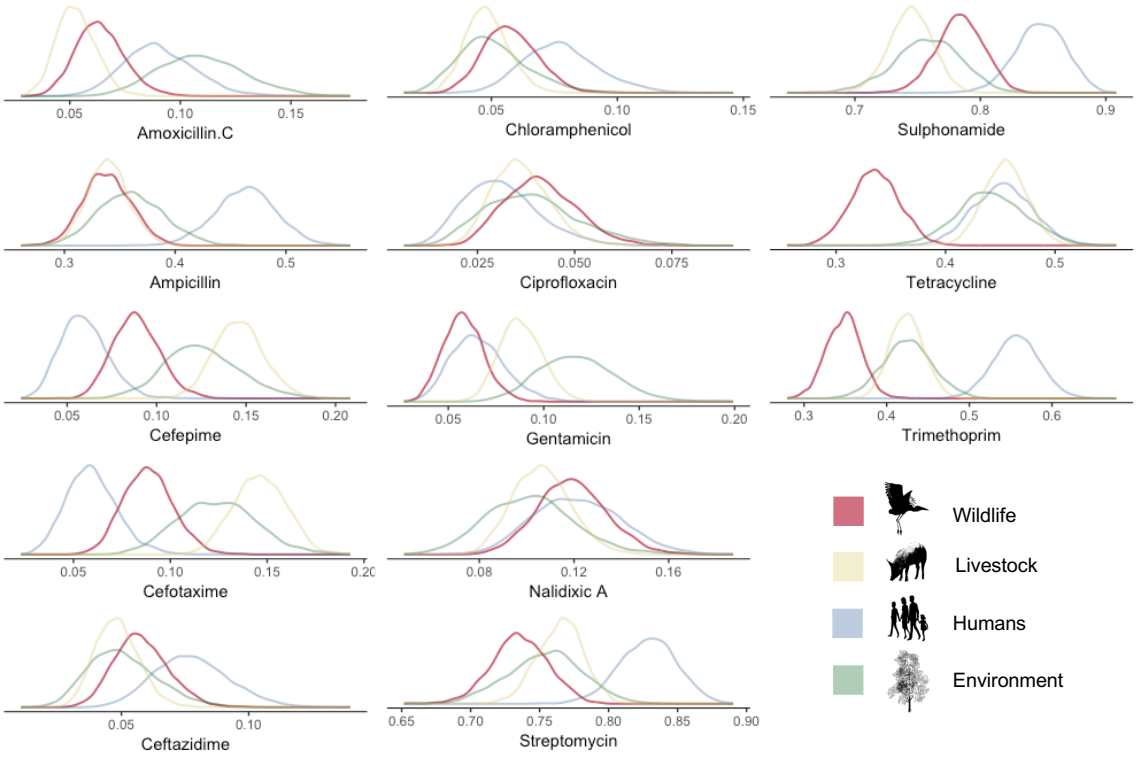


Figure 5.2. Bayesian posterior distributions for the prevalence (x-axis) of 13 antibiotics in wildlife (red), livestock (green), human (blue) and environmental (yellow) samples.

5.4.1 Differentiation between epidemiological compartments in Nairobi

To determine whether urban wildlife in Nairobi are a net recipient of AMR from other epidemiologically linked compartments (wildlife, livestock, humans and the environment), prevalence, MDR, profile length and diversity of phenotypes were compared between compartments. Prevalence of AMR-*E. coli* in wildlife was significantly lower than one or more other compartments for six of the antibiotics tested (ampicillin, cefepime, cefotaxime, streptomycin, tetracycline and trimethoprim), and not significantly higher than other compartments for any of the 13 antibiotics tested (Figure 5.2, Suppl. Table 5.3). Variation in the likelihood of MDR-*E. coli* carriage, and variation in total profile length, between compartments was modelled using binomial (with log-link function) and Poisson GLMMs respectively. Intercept-only mixed-effects models fitted to the presence/absence of MDR and total profile length showed that the household in which samples were collected from accounted for a larger proportion of total variation (21% and 3% of variation in each model, respectively). For both models, fit was improved by dropping sublocation, and as such only the household dependency structure was carried forward to the full models. Distance-based Moran's eigenvector maps (dbMEMs) were used to explore the spatial structure of the dataset, and a total of 13 dbMEM eigenvectors modelled positive spatial correlation between all samples. Five eigenvectors (describing broad, medium and fine-scale spatial structure) were significantly associated with MDR-*E. coli* carriage and AMR profile length of isolates, and thus included as covariates in both full models. Step-wise selection resulted in the inclusion of the fixed covariate source compartment (categorical with four levels), and three spatial eigenvectors (continuous), leading to models of the form

$$\mathrm{MDR}_{ij} = \text{Source Compartment} + \mathrm{MEM1}_{ij} + \mathrm{MEM2}_{ij} + \mathrm{MEM5}_{ij} + \mathrm{Household}_{i} \quad \text{(eqn 1)}$$

$$\text{Profile Length}_{ij} = \text{Source Compartment}_{ij} + \mathrm{MEM1}_{ij} + \mathrm{MEM2}_{ij} + \mathrm{MEM5}_{ij} + \mathrm{Household}_{i} \quad \text{(eqn 2)}$$

where MDR_{ij} and $Profile\ Length_{ij}$ are the jth observation in household i, and $i = 1,...,99$, and $household_i$ is the random intercept, which is assumed to be normally distributed with mean 0 and variance σ^2. Validation procedures were conducted for the models in eqns 1 and 2 as described in the methods, and did not indicate any problems. Wildlife were significantly less likely to carry MDR-*E. coli* than humans and livestock, and had significantly shorter profile lengths than all other compartments (Suppl. Table 5.4). Both models also showed significant broad-scale spatial relationships for AMR carriage across the city; the probability of MDR-*E.coli* carriage in all epidemiological compartments decreased along an East – West gradient (MEM1), whilst AMR profile lengths decreased from Eastern – Western Nairobi (MEM1) and increased from Northern – Southern Nairobi (MEM2). The models accounted for 2.8% (eqn 1) and 4.3% (eqn 2) of variation in the data (marginal R^2: 0.028, 0.043).

Population-diversity measures of resistance indicated that wildlife had less diverse AMR profiles than other compartments. *E. coli* isolated from wildlife had a lower expected AMR profile diversity than all other compartments as measured by three of the four D_α diversity indices calculated [Shannon entropy (SEn), Simpson diversity (SD) and Berger-Parker (BP)] (Figure 5.3, Suppl. Table 5.5). When compared across all compartments, the range of median Alpha values was statistically significantly lower in wildlife than all other compartments (Wildlife:Environmental, $p<0.01$; Wildlife:Livestock, $p<0.01$; Wildlife:Human, $p<0.001$). Asymptotic AMR profile richness was estimated using the Chao2, ICE and Jacknife statistical models, and when all 262 AMR profiles present in the dataset were considered, accumulative estimates for each compartment began to show signs of saturation (Figure 5.4, Suppl. Table 5.6). Asymptotic estimates of total AMR profile richness in wildlife were 273 (95% CI 245-300) unique profiles, all of which could be detected if an additional 8848 extra samples had been collected. This richness estimate is lower than estimates for the environment and livestock (350 [95% CI 305-395] and 416 [95% CI 378-454] respectively), but higher than humans (185 [95% CI 165-205]). Unlike the human compartment, where estimates stabilised at 270 samples (suggesting that most of the predicted profile diversity had been detected), wildlife and livestock estimates were only beginning to stabilise at the sampling extremes achieved in this project. None of the statistical estimators stabilised for environment, suggesting that the rate of discovery of new profiles in this compartment was still high, and minimum richness estimates could therefore be considerably higher than 350.

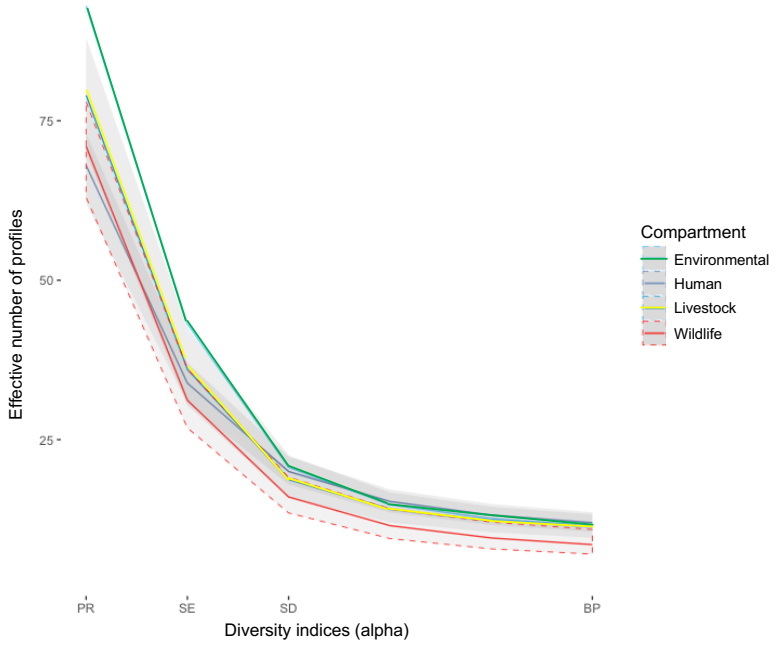


Figure 5.3. Diversity of AMR profiles (measured as the effective number of AMR profiles) across four diversity indices, for environmental, human, livestock and wildlife compartments. Shaded areas (grey) represent 95% CI intervals around the median for each line. Dashed line (red) highlights 95% CI for wildlife. PR, profile richness; SE, Shannon entropy; SD, Simpson diversity; BP, Berger-Parker. Ranges for Alpha were significantly lower in wildlife than for all other compartments.

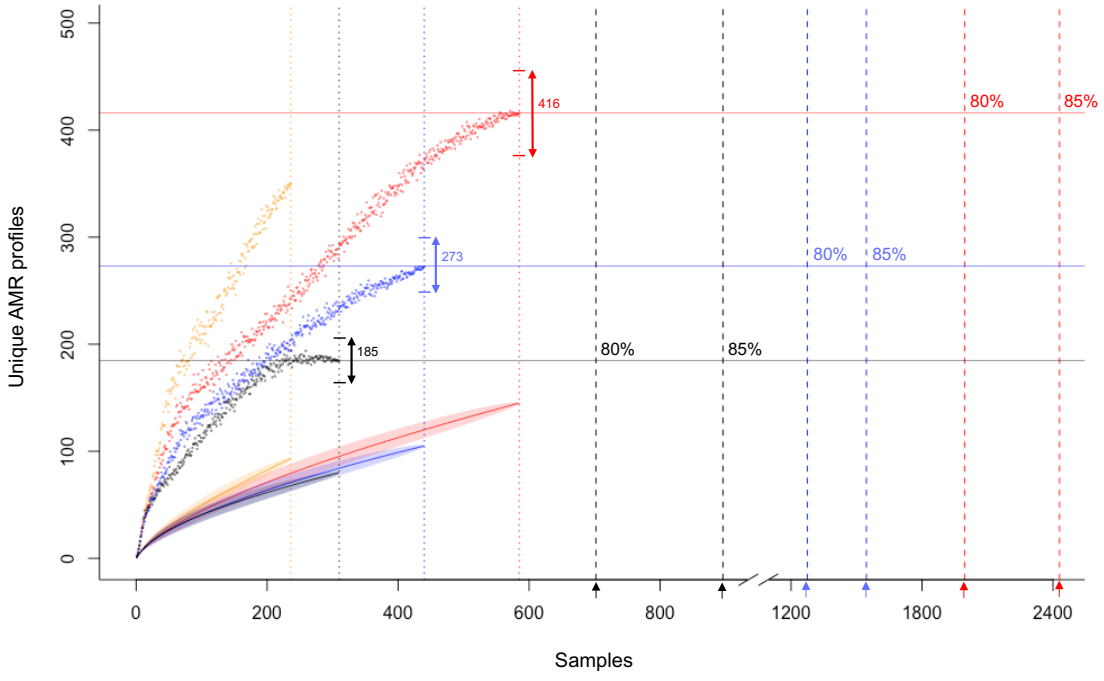


Figure 5.4. Asymptotic AMR profile richness estimates for each compartment (Wildlife, Blue; Livestock, Red; Human, Black; Environmental, Yellow). Dotted curves, Chao 2 estimators at every sample point (95% CIs indicated by double-ended arrows at asymptote); Horizontal lines, asymptotic estimate of profile richness for each compartment; Shaded curves, species accumulation curves (line represents model fitted values, shaded areas represent 95% CIs); Vertical dotted lines (…), number of samples collected from each compartment in this study; Vertical dashed lines (- - -), sampling effort required to detect 80% and 85% of the asymptotic estimate for AMR profile richness in each compartment.

5.4.2 The role of wildlife functional ecology in AMR-*E.coli* carriage

When split into taxonomic groups, prevalence of resistance to at least one antibiotic was 90.7% (440/485) across all wildlife, and 90.8% (256/282), 90% (18/20), 91.6% (142/155) and 85.7% (24/28) in birds, avian populations (APs), rodents and bats, respectively. *E. coli* resistant to each of the 13 antibiotics tested were detected in all taxonomic groups of wildlife except bats, and Bayesian models showed that prevalence of resistance to streptomycin, tetracycline and trimethoprim varied significantly between wildlife when stratified by taxonomic and/or functional groups (Figure 5.5, Suppl. Table 5.3). Frugivorous birds and bats were the only functional group for which *E. coli* resistant to each of the 13 antibiotics were not detected. Wading birds belonging to the orders *Pelecaniformes* and *Ciconiiformes* were more likely to carry *E. coli* resistant to ceftazidime (OR: 7.9; 95% CI 1.7-28.5; $p<0.01$), and had significantly longer profiles than other species of wildlife ($p<0.05$).

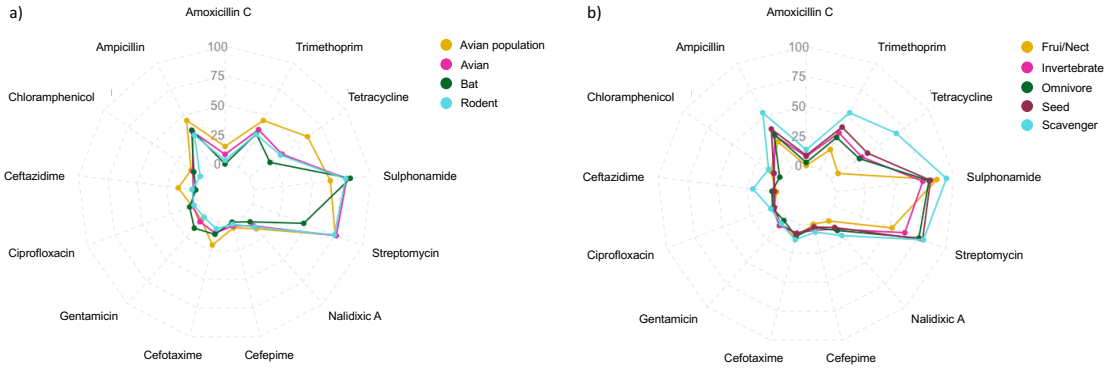


Figure 5.5. Radar plots showing differences in prevalence for *E. coli* resistance to the 13 antibiotics tested for between **a)** wildlife taxonomic groups (avian populations, birds, bats and rodents); **b)** wildlife function groups stratified by feeding ecology.

To model the likelihood of wildlife carrying MDR-*E. coli* as a function of feeding ecology, a Binomial GLMM with a log-link function was used [eqn 3]. An intercept-only mixed-effects model fitted to the presence/absence of MDR in wildlife showed that the household and sublocation in which animals were sampled accounted for a small proportion of total variation (4.3% and 5.3% of variation respectively). Model fit was improved by dropping household, and as such only the sublocation dependency structure was carried forward to the full model. Distance-based Moran's

eigenvector maps (dbMEMs) were used to explore the spatial structure of the dataset, and a total of 32 dbMEM eigenvectors modelled positive spatial correlation. However, only a single eigenvector (describing very broad, longitudinal spatial structure between Eastern and Western Nairobi) was significantly associated with MDR-*E. coli* carriage ($p<0.0001$), and thus included as a covariate in the full model (Supple. Figure 4.2). No evidence of a temporal relationship for MDR-*E. coli* carriage was found, and as such the week in which animals were sampled was not included in the full model (Suppl. Figure 5.4). Step-wise selection resulted in the inclusion of the fixed covariates *taxonomic functional group* (categorical with eight levels), and the spatial eigenvector *MEM1* (continuous), leading to a model of the form

$$\mathrm{MDR}_{ij} = \mathrm{TaxonFunc}_{ij} + \mathrm{MEM1}_{ij} + \mathrm{Sublocation}_{i} \quad \text{(eqn 3)}$$

where MDR_{ij} is the jth observation in sublocation i, $i = 1,...,33$, and $\mathrm{sublocation}_i$ is the random intercept, which is assumed to be normally distributed with mean 0 and variance σ^2. Validation procedures for the model in eqn 3 indicated no patterns in the residuals, or evidence of spatial autocorrelation. MDR-*E. coli* carriage varied by taxonomic functional groups, and along an East-West gradient across Nairobi, although the fixed covariates explained little of the total variation in the dataset (marginal R^2: 0.08). Frugivorous bats, and seed-eating, omnivorous and scavenging birds were significantly more likely to carry MDR-*E. coli* than frugivorous birds, and the probability of carrying MDR-*E. coli* showed a decreasing trend from East to West Nairobi (Figure 5.6, Table 5.7).

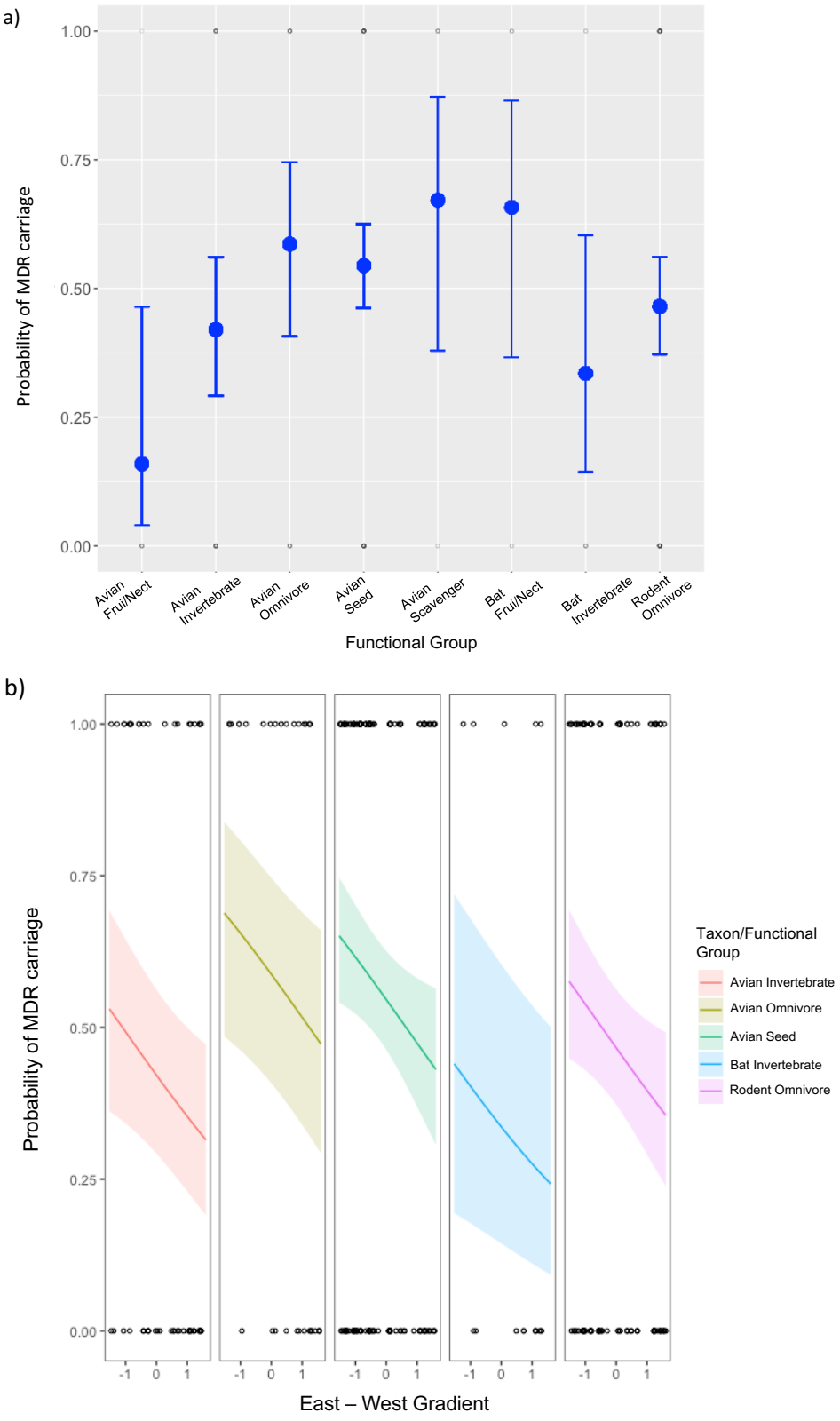


Figure 5.6. Fit of the Binomial generalised linear mixed effects model (GLMM) shown in eqn (3), relating MDR-*E. coli* carriage in wildlife to their feeding ecology. Graphs show variation in likelihood of MDR-*E.coli* carriage for **a)** different wildlife taxonomic/functional groups; **b)** the East – West broad spatial gradient across Nairobi, further stratified by wildlife taxonomic/functional group. Bands in **a)** and coloured shading in **b)** represent 95% confidence intervals.

Significant variation in *E. coli* AMR profile length was also found when modelled against covariates in the form of a Poisson GLMM. Step-wise selection against a random-intercept model resulted in household and sublocation dependency structures being fitted as random effects, and six dbMEM eigenvector functions being included as fixed effects, along with taxonomic group (categorical with three levels) and laboratory (categorical with two levels) (Suppl. Figure 5.2). This model took the form

$$\begin{aligned}\mathrm{MDR}_{ijk} = \mathrm{Taxon}_{ijk} + \mathrm{MEM1}_{ijk} + \mathrm{MEM8}_{ijk} + \mathrm{MEM10}_{ijk} + \mathrm{MEM19}_{ijk} + \mathrm{MEM25}_{ijk} + \mathrm{MEM27}_{ijk} + \\ \mathrm{Laboratory}_{ijk} + \mathrm{Household}_{i} + \mathrm{Sublocation}_{j}\end{aligned}$$

(eqn 4)

where MDR_{ijk} is the kth observation in household i, sublocation j (i = 1,...,99, j = *1,...,*33), and $\mathrm{household}_i$ and $\mathrm{sublocation}_j$ are random intercepts, which are assumed to be normally distributed with mean 0 and variance σ^2. For the model in eqn 4, *E. coli* AMR profiles were longer in birds than rodents, and profile length showed spatial correlation across multiple scales of the city [broad-scale (East-West; MEM1), medium-scale (MEM8, 10, 19) and fine-scale (MEM25, MEM27) resolutions] (marginal R^2: 0.13) (Table 5.7). Longer AMR profiles were more likely to be identified in wildlife samples tested at the University of Nairobi.

5.4.3 AMR exchange between wildlife, livestock and humans within households

Having established that wildlife are a more probable 'sink' rather than source population for AMR *E. coli* in Nairobi (being characterised by lower MDR-*E.coli* carriage, shorter AMR profiles, and less phenotypic diversity when compared to other compartments), household interfaces were assessed for AMR exchange between wildlife, livestock and humans. Seed-eating birds and rodents are ubiquitous in households across Nairobi, frequently display anthropophilic feeding behaviour, and have already been documented in this study as potentially important carriers of MDR-*E. coli*. As such, AMR in these functional groups was used as the basis of efforts to understand epidemiological connectivity within households. To investigate potential

interfaces for AMR exchange within households, ecological models were developed to predict the likelihood of MDR-*E. coli* carriage in seed-eating birds and rodents as a function of household-level anthropogenic and ecological covariates (Suppl. Figure 5.3, Suppl. Table 5.2). For MDR carriage in seed-eating birds, the best-fitting multivariate model includes the fixed covariates *total cattle* (count, log-transformed), whether *garbage* and *manure* are disposed of within the household compound or outside (both categorical with two levels), and *median AMR profile length* for *E. coli* isolated from *livestock* and *humans* (both continuous). The interaction terms are *median livestock AMR profile length* × *manure disposal*, and *median human AMR profile length* × *garbage disposal*. A random-intercept model indicated zero dependency at the household or sublocation level, and as such a random intercept was not included.

$$\begin{aligned}\mathrm{MDR}_i = {} & \mathrm{Total\ Cattle}_i + \mathrm{Garbage}_i + \mathrm{Manure}_i + \mathrm{Livestock\ profile\ length}_i + \mathrm{Human\ profile\ length}_i \\ & + \mathrm{Livestock\ profile\ length}_i \times \mathrm{Manure}_i + \mathrm{Human\ profile\ length}_i \times \mathrm{Garbage}_i\end{aligned}$$

(eqn 5)

For MDR carriage in rodents, the best fitting model (depicted in eqn 6) included whether *garbage* and *manure* are disposed of within the household compound or outside (both categorical with two levels), *median AMR profile length* for *E. coli* isolated from livestock and humans (both continuous), and *Laboratory* (categorical with two levels). The interaction terms were *median human AMR profile length* × *manure disposal*, and *median human AMR profile length* × *garbage disposal*. As above, a random-intercept model indicated zero dependency at the household or sublocation level, and as such a random intercept was not included in the model.

$$\begin{aligned}\mathrm{MDR}_i = {} & \mathrm{Garbage}_i + \mathrm{Manure}_i + \mathrm{Livestock\ ABG}_i + \mathrm{Human\ ABG}_i + \mathrm{Laboratory}_i + \mathrm{Human\ ABG}_i \times \mathrm{Garbage}_i \\ & + \mathrm{Human\ ABG}_i \times \mathrm{Garbage}_i\end{aligned}$$

(eqn 6)

Model validation indicated no patterns in the residuals, or evidence of spatial autocorrelation for either model. The avian model in eqn 5 demonstrates that in any given household, the likelihood of MDR carriage in seed-eating birds is best described by increasing numbers of cattle in the household compound, and AMR profile length

of the human inhabitants (R^2: 0.3; Figures 5.7 and 5.8, Table 5.8). The relationship between human profile length and avian MDR carriage is influenced by whether garbage was kept within the household compound or not; keeping garbage within the compound results in a stronger relationship between human profile length and avian MDR carriage. When manure is kept inside the household compound, the probability of MDR carriage in seed-eating birds increases with longer AMR profile lengths in livestock, whilst the opposite is true when manure is disposed of externally (Figure 5.8). For the rodent model in eqn 6, the likelihood of MDR carriage increases with increasing profile length of livestock and human inhabitants in the household (R^2: 0.42; Figure 5.9, Table 5.8). Although not statistically significant within the model, keeping both garbage and manure outside the house reduces the likelihood of rodents carrying MDR as human profile length increases.

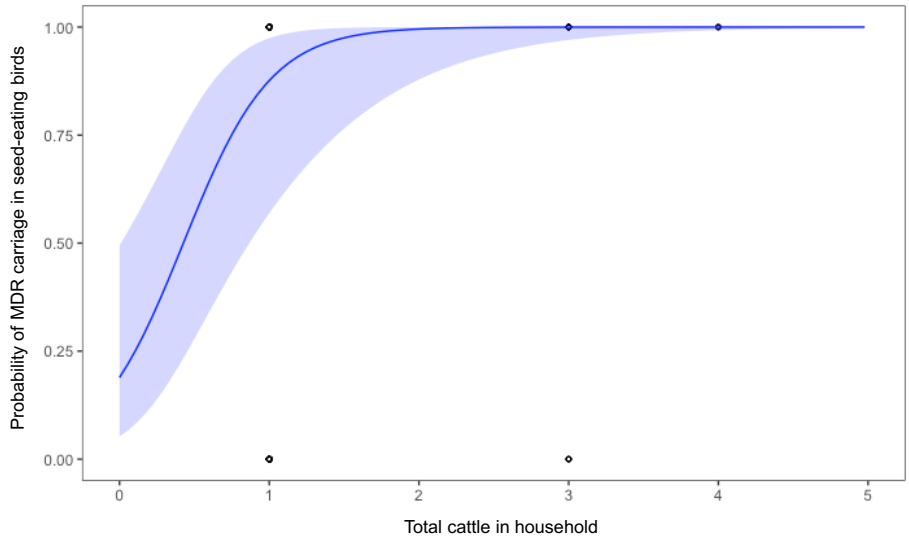


Figure 5.7. Fit of the Binomial generalised linear mixed effects model (GLMM) in eqn (5), relating multidrug resistant (MDR)-*E. coli* carriage in birds to household-level anthropogenic and ecological covariates. This graph shows how total cattle in a household influences likelihood of MDR-*E. coli* carriage in seed-eating birds. All other covariates in the model are kept constant. Blue shading represents 95% confidence intervals along the regression line, and black points indicate data points.

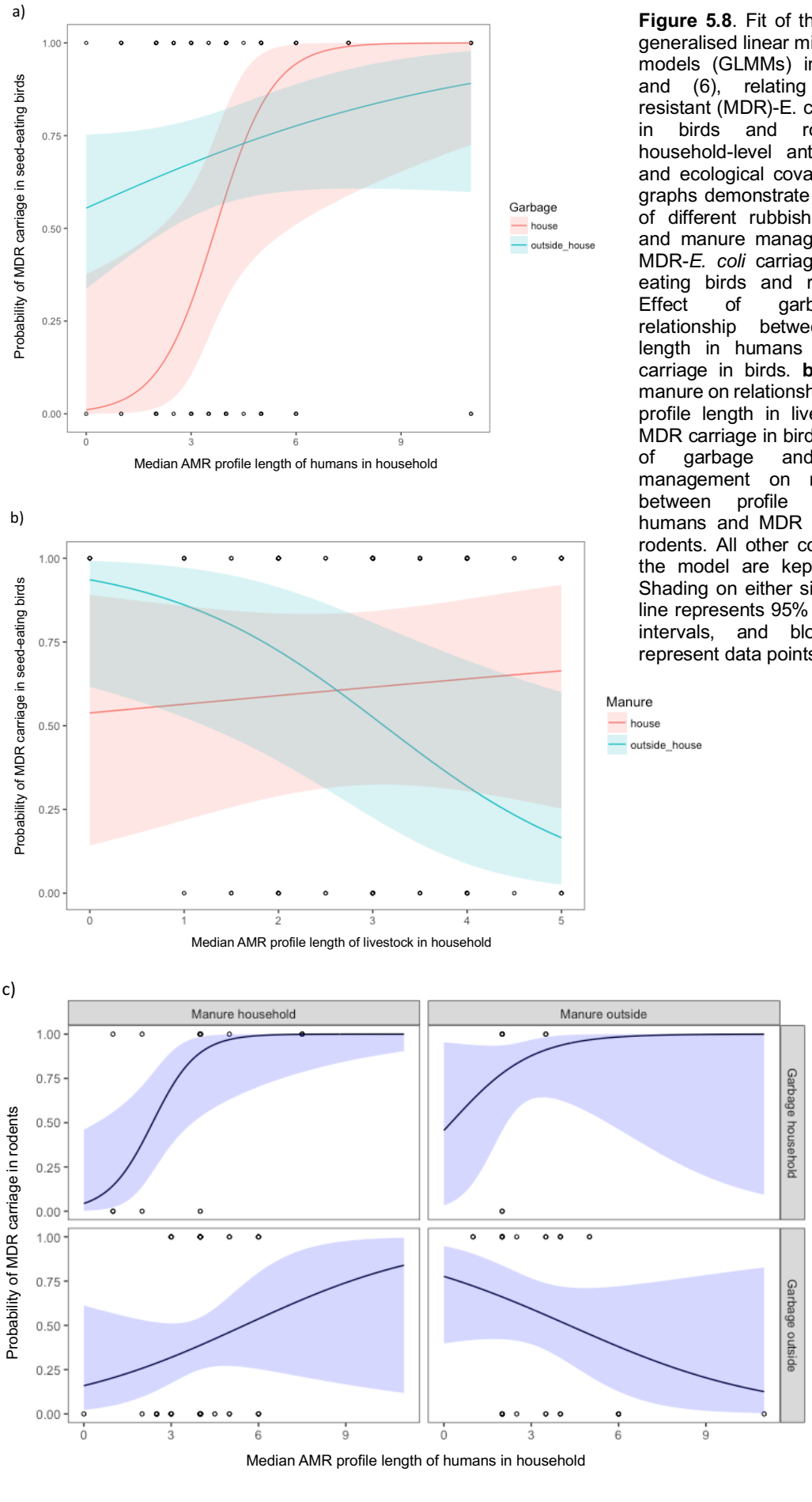


Figure 5.8. Fit of the Binomial generalised linear mixed effects models (GLMMs) in eqns (5) and (6), relating multidrug resistant (MDR)-E. coli carriage in birds and rodents to household-level anthropogenic and ecological covariates. The graphs demonstrate the effects of different rubbish (garbage) and manure managements on MDR-*E. coli* carriage in seed-eating birds and rodents. **a)** Effect of garbage on relationship between profile length in humans and MDR carriage in birds. **b)** Effect of manure on relationship between profile length in livestock and MDR carriage in birds. **c)** Effect of garbage and manure management on relationship between profile length in humans and MDR carriage in rodents. All other covariates in the model are kept constant. Shading on either side of each line represents 95% confidence intervals, and block points represent data points.

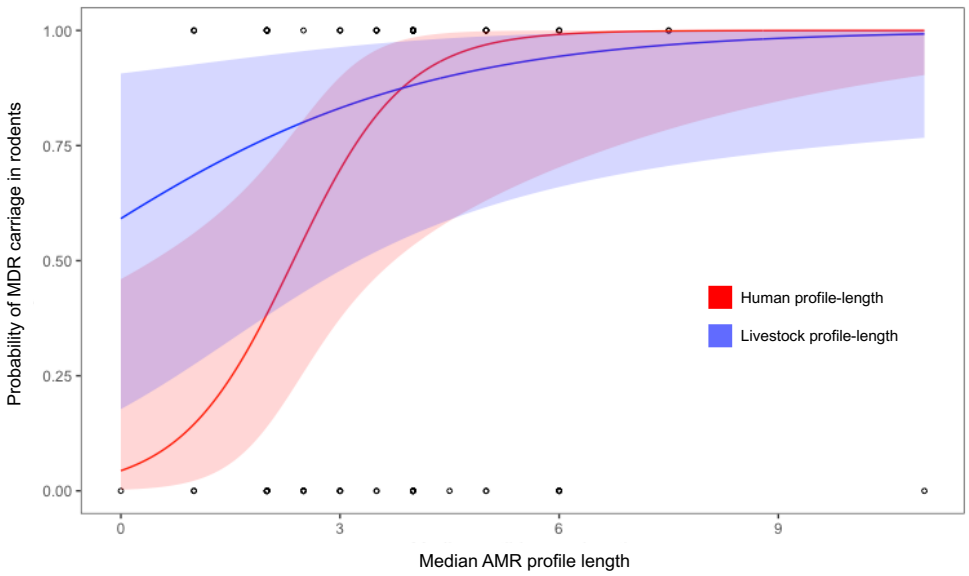


Figure 5.9. Fit of the Binomial generalised linear mixed effects model (GLMM) in eqn (6), relating multidrug resistant (MDR)-E. coli carriage in rodents to household-level anthropogenic and ecological covariates. This graph shows how increasing human and livestock profile lengths in a household influences the probability of MDR-*E. coli* carriage in rodents. All other covariates in the model are kept constant. Shading on either side of each line represent 95% confidence intervals, and black points indicate data points.

5.5 Discussion

The environment could act as an important reservoir for antimicrobial resistance, but little is known about the ecology of resistant genes outside human and livestock hosts. Wildlife exist across multiple trophic levels, and are therefore well placed to accumulate and disperse resistance determinants within ecosystems. In this study, the role that urban-adapted wildlife play in the epidemiology of antimicrobial resistance in Nairobi is addressed. Unlike most previous studies on AMR in wildlife, in which wild animals have been opportunistically sampled [121], this study employed an epidemiological study design to compile a large bacterial dataset for investigating AMR exchange between sympatric wildlife, humans, livestock and their shared environment. The findings from this study show that carriage of clinically relevant AMR phenotypes in urban wildlife is widespread, and dictated by epidemiological exposure through habitat utilisation.

5.5.1 Epidemiological role of wildlife AMR in Nairobi

At the broadest level, each epidemiological compartment in Nairobi (wildlife, livestock, humans and the environment) either acts as a net source, or sink for clinically relevant AMR-bacteria. In urban environments, the latter seems more probable for wildlife species that are closely integrated into anthropogenic habitats. This hypothesis was tested using measures of ecological diversity that consider different weightings of species richness and abundance in relation to diversity. For most of these measures, AMR profiles were significantly less diverse in the wildlife population than the environmental, human and livestock compartments, which is consistent with the hypothesis that wildlife are not a major source of AMR diversity in Nairobi, and that diversity is generated elsewhere. Additional differences between MDR-*E. coli* carriage and AMR profile length between epidemiological compartments support this assertion. Non-parametric species discovery curves, which have been used to estimate viral diversity in mammals [184], were also used to make statistical estimates for the fraction of undetected AMR diversity (measured as profile richness) in each compartment. Acknowledging that these values represent a minimum estimate of richness, total profile richness in wildlife was found to be considerably lower than livestock and the environment, adding further strength to our findings. Interestingly, estimates for total profile richness in humans was considerably lower than all other compartments. This difference, which was robustly supported by statistical estimators showing reliable asymptotic behaviour for AMR profile discovery in human samples, might indicate that, compared to humans, wildlife (and potentially livestock) are exposed to greater AMR diversity through their closer interaction with the environment. Although these results suggest that the net source of AMR diversity lies within other compartments, this is not to say that AMR never arises in wildlife. Exposure to external and pre-existing enteric AMR genes could lead to selection processes in enteric bacteria, which are subsequently excreted into the environment [121].

The vertebrate microbiome is likely to play a key role in the population structure of enteric AMR genes, and microbiome composition is dictated by an array of factors linked to host genotype, life-stage and diet [192]. Recent studies have

demonstrated that wildlife with no obvious exposure to antibiotics carry diverse resistant profiles that vary by host and season [193]. That significant variation in MDR carriage and profile length were found according to taxonomic and feeding niches is therefore unsurprising, and is in line with other studies that have reported higher carriage of AMR bacteria in carnivores and omnivores [171]. However, without culture-independent characterization of microbial communities, separating the direct effects of diet and physiological factors on selection for enteric AMR genes, and foraging-related life history traits that could increase exposure to external sources of AMR determinants, is very difficult. For example, at the taxon level in this study, birds tended to carry *E. coli* with more resistant phenotypes than rodents, but it is unclear to what extent this is due to physiological differences between birds and mammals, variation in exposure to external sources, or a combination of the two. Although pre-existing carriage or within-host factors could not be assessed, compelling evidence was found to suggest that in certain species of urban wildlife *E. coli* resistance to clinically relevant antibiotics is influenced by foraging behaviour. Scavenging birds (such as Marabou Storks), which occupy high trophic levels and as such have large home ranges and foraging behaviours that associate them with human and animal waste, were more likely to carry resistance to Tetracycline and Trimethoprim antibiotics, and multidrug resistance, than functional groups with more localised feeding habits (such as frugivores and nectarivores). Colony forming birds belonging to the orders *Pelecaniformes* and *Ciconiiformes* also had longer AMR profiles and a higher prevalence of resistance (particularly to third and fourth generation cephalosporins). In the absence of natural habitats such as wetlands, these species forage on sewage treatment plants, rubbish dumps and abattoir viscera ponds. Artificial habitats such as these are considered important routes for the dispersal of human and livestock excreted AMR into the environment [65,121].

Having established that wildlife in Nairobi are a net ‘sink’ for clinically relevant AMR phenotypes, and that exposure could be determined by foraging behaviour, evidence for exchange between wildlife and livestock, humans or the environment within household compounds was assessed. To investigate this, modelling was applied to multidrug resistance in anthropophilic seed-eating birds and rodents that were trapped within livestock keeping houses across Nairobi. For both seed-eating birds and rodents, multivariable mixed models provide evidence for the existence of two ‘interfaces’ within household compounds. Firstly, an interface for the

exchange of AMR-*E. coli* exists between seed-eating birds, cattle and bovine manure. The likelihood of AMR carriage in seed-eating birds was linked to the presence of livestock (particularly cattle) in household compounds, which was coupled with a 'manure' effect; if manure was kept inside the household compound, the likelihood of MDR carriage in birds increased with the AMR profile length of livestock-*E.coli*. Evidence of a similar interface between livestock and rodents was found, although without the involvement of manure. A second interface exists between humans, their rubbish and seed-eating birds. The median length of human AMR profiles in each house was linked to higher carriage of MDR in seed-eating birds, and this relationship was stronger when rubbish was kept within the household compound. Evidence of a similar human-rubbish interface for rodents was present, where the likelihood of MDR carriage in a household increased with median profile length for human inhabitants. Although non-significant in the model, keeping garbage and manure inside the household compound appears to increase the likelihood of exchange across this interface. Seed-eating birds (such as House Sparrows) were frequently observed interacting closely with manure and rubbish during data collection, and these results suggest that anthropogenic resource provision within households provides a direct route of phenotypic AMR exchange between wildlife and livestock, and wildlife and humans. However, whilst these results are suggestive of AMR exchange across these interfaces, genetic data is necessary to corroborate their existence, and determine whether bacteria are being transferred across them.

5.5.2 Implications for human health, ecosystem health and surveillance

Wildlife borne AMR probably poses little direct threat to human health in urban areas. Although high levels of *E. coli* resistant to clinically relevant antibiotics were detected in urban wildlife, including resistance to the more newly developed, synthetic drugs such as fluoroquinolones, direct contact between humans and wildlife is rare. Livestock and environmental compartments (which both had higher ecological diversity of AMR and with which humans have more direct contact) are more likely routes of human exposure to novel AMR genes, although assessing the risk posed by these compartments lies far beyond the scope of this study [194–196]. Instead, high levels of resistance and MDR-*E. coli* carriage, and the presence of XDR and PDR in wildlife, are probably indicative of environmental antibiotic contamination in the city.

It has been suggested that wildlife, particularly rodents and migratory birds, could act as 'sentinels' for AMR, being used to inform of the public health risk posed by environmental pathogen pollution [173,197]. If exposure to AMR in wildlife is largely determined by habitat usage, targeted surveillance of wildlife that frequent 'high risk' urban environmental interfaces (where the accumulation of clinical residues in high concentrations might force the accelerated evolution and fixing of resistance determinants) could represent an efficient way to detect clinically important resistance determinants. For example, population-level samples of scavenging and wading birds are technically simple to collect from roosting sites, and could represent an accumulation of AMR genes acquired from multiple habitats. To explore the practicality of conducting surveillance in wildlife, the sampling effort required to detect different fractions of the total estimated AMR richness was calculated (Figure 5.4, Table 5.6). To detect all 273 predicted AMR profiles in the wildlife species sampled would require an extra 8848 samples, an impractical and expensive task. However, detecting 85% of the total diversity would require a disproportionately lower sampling effort of 1572 samples. Assuming that the diversity of resistance profiles in selected scavenging and wading birds is lower than the total diversity represented by all taxonomic classes of wildlife included in this study, the required sampling effort to achieve an acceptable likelihood of detecting new profiles in these species would be much lower. Extending this approach to livestock and humans reveals similar outcomes for surveillance of AMR profiles in these compartments (Figure 5.4). If changes in phenotypic and genotypic diversity mirror one another, practical and economically viable surveillance for AMR genes of public health concern in urban wildlife, livestock and humans could be achieved through targeted surveillance, designed to capture a high proportion of diversity.

Developing country urban ecosystems with high levels of background environmental AMR could act as pools of AMR dissemination to peripheral ecosystems, where the flow of water, and movement of humans, livestock and wildlife act as vectors for dispersal. Little is known about how resistance genes are carried and shed by wildlife species [121]. Wildlife with larger home ranges are more likely to travel beyond city limits, and the finding of high AMR carriage in scavenging birds (which typically have large home ranges), suggest that these species could disseminate AMR determinants to neighbouring ecosystems - Nairobi is surrounded by a complex patchwork of high density human populations, natural areas, forest and rangelands.

Mapping the distribution of MDR-*E.coli* in wildlife by sublocation shows high levels of MDR carriage extending to peripheral areas of Nairobi, which border rich Savannah ecosystems to the South and East of the city (Figure 5.10). Nairobi National Park, which borders the city to the South, is home to a high density of migratory plains species that could disperse AMR genes to more distant areas [198]. Within the city, models for MDR-*E.coli* carriage and profile length showed a clear East-West gradient, suggesting that wildlife AMR diversity is higher in the East of the city. This corresponds with the extreme environmental, ecological and social gradients that split the city; the West is wealthy, forested, ecologically diverse and at a higher altitude, and the East is poorer, more urbanised (and thus ecologically homogenous) and at a lower altitude. Soil type and the flow of water also follow the topographical gradient of the city, with rivers crossing from West to East. Such extreme differentiation within a single city demonstrates the highly complex ecosystem within which the epidemiology of AMR is set.

From a public health perspective, the importance of these results extends beyond AMR. Urban resource provisioning has been identified as an important driver of changes to wildlife population dynamics, and a factor that can bring wildlife into closer association with humans and livestock, thus offering new opportunities for disease emergence [31,90,199]. Although the household-level interfaces identified in this study are likely to pose little risk to humans and livestock through exchange of AMR determinants with wildlife, their existence represents important signals of 'epidemiological connectivity' within households; interfaces across which other microorganisms, such as bacteria and viruses, could be exchanged. More research, preferably utilising next-generation sequencing data, is required to define the structure of these interfaces and transfer of microorganisms across them. Whilst we caution against trialling interventions before interfaces have been properly defined, our results suggest that action could be taken to disrupt transmission at household wildlife-livestock-human interfaces. Removing manure and rubbish (sources of anthropogenic resource provision) from households appeared to reduce the magnitude of AMR exposure in birds and rodents, either through limiting wildlife-livestock/human contact or reduced exposure of wildlife to sources of AMR.

I employed a similar approach to assessing AMR diversity in this study as Mather et al. [1], and advocate the use of phenotypic rather than genotypic data along comparable lines. Given that the objective of this study was to describe the role of

wildlife in the epidemiology of clinically relevant AMR, using phenotypic data makes these results more accessible to clinicians, veterinarians and policy-makers, and

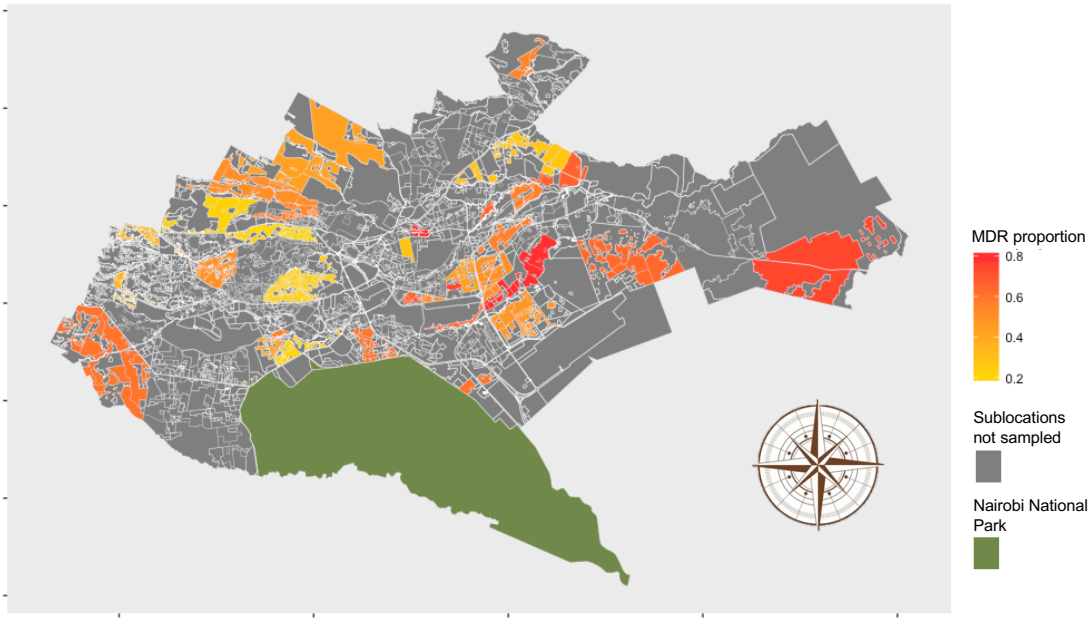


Figure 5.10. Proportion of multidrug resistant-*E coli* carriage in wildlife, stratified by the sublocation in which they were sampled.

comparable to studies in other locations where genotypic expression of AMR-*E.coli* might vary. To this effect, using internationally recognised breakpoint methods and definitions for MDR, XDR and PDR aligns this analysis as closely as possible with clinical standards for AMR surveillance in humans. However, molecular data would be a valuable addition to this dataset, and could be used to support, strengthen or refute our assertions. Sequence data would be particularly valuable in examining the connectivity and exchange of AMR genes at interfaces, and metagenomics could be used to study within-host AMR diversity. Comparisons between phenotypic and genotypic expression for *E. coli* belonging to the same dataset are currently being conducted using whole genome sequencing (WGS), and should indicate how closely the phenotypic and genetic predictions map together. The fine scale analysis permitted by WGS might also shed light on whether within-host selective pressure in wildlife promotes the accumulation of novel AMR genes, that could be excreted into the environment. Samples were tested at one of two laboratories in Nairobi, and, for models including wildlife samples, was treated as a confounding factor (laboratory was not included in models including all epidemiological compartments, because samples from different sources were unevenly split between labs). In two models,

laboratory influenced the detection of AMR in samples. Protocols were standardised between laboratories, so it is likely that this variation arose through operator bias. However, given that the laboratory effect was only present in two models and had opposing effects on MDR carriage in each, it can be considered a limited rather than systematic source of error, that is unlikely to have had a large effect on this data.

Whilst the prevalence of resistance to individual antibiotics and MDR-*E. coli* carriage detected in wildlife in this study was high, without comparable datasets from other urban or rural settings it is difficult to say how unique these results are to Nairobi. Similarly scaled studies conducted in a variety of urban settings (examples of which I am not aware of) would permit examination of context-specific differences in wildlife AMR carriage and exposure. The effort required to sample wildlife is high, and varies greatly between species. As such, the sample size in this study was relatively small for 'cryptic' taxonomic and functional groups (e.g. bats, scavengers and frugivores/nectarivores), which introduced a level of uncertainty to the prevalence estimates, statistical tests and ecological models in which these species were considered. Given the relative importance of scavenging and wading birds, more effort should be devoted to sampling these species in future. In addition, our focus on mammalian and avian urban wildlife neglects the role played by reptiles, aquatic organisms and invertebrates. Studies investigating the effects of AMR on invertebrates, and their role in carriage and dispersal of resistance elements are warranted given the indispensable role invertebrates play as pollinators, bio-control agents, and in the degradation and recycling of organic matter in soils [200,201].

5.6 Conclusion

In this chapter, ecological and epidemiological approaches were applied to provide the first comparative analysis of phenotypic AMR in sympatric wildlife, livestock, humans and the environment in an urban setting. A high prevalence of *E. coli* resistant to clinically relevant antibiotics were found in wildlife, with a significant proportion of animals demonstrating multidrug resistance. Wildlife are an unlikely source of clinically relevant resistance in Nairobi, with exposure being determined by feeding ecology, and indirect exchange of AMR occurring at household interfaces

between seed-eating birds, rodents, livestock and humans. These results provide insight into the determinants of AMR carriage in wildlife, and routes of exposure between wildlife, livestock and humans. Ultimately, this work in Nairobi forms part of a broader strategy to understand the epidemiology of AMR, and identify routes of epidemiological connectivity, across developing urban landscapes.

5.8 Supplementary Tables

Antibiotic Class	Antibiotic Drug
Aminoglycoside	Gentamicin
	Streptomycin
β-lactam	Amoxicillin-clavulonic Acid
Cephalosporin	Cefepime
	Cefotaxime
	Ceftazidime
Chloramphenicol	Chloramphenicol
Folic Acid	Trimethoprim
Penecillin	Ampicillin
Quinolone/Fuoroquinolone	Cirpofloxacin
	Nalidixic Acid
Sulfonamide	Sulfamethoxazole
Tetracycline	Tetracycline

Table 5.1. Antibiotic drugs to which sensitivity of *E. coli* isolates was tested for in this study, and the antibiotic classes to which these drug belong.

Covariate		Format	Project source	External source	Notes
Ecological	Taxonomic group	Categorical (Avian popultaion, Avian, Bat, Rodent)	Metadata		
	Feeding niche	Categorical (Fruit/Nectar, Omnivore, Invertebrate, Seeds, Scavenger)		[43]	
	Epidemiological compartment	Categorical (Wildlife, Environment, Livestock, Humans)	Metadata		
	Proportion of time spent foraging on ground	Continuous (Proportion)		[43]	
	Home range	Continuous (Hectares) (Log[10])	Metadata (body weight)	[45-48]	Not calculated for bats
	Association of species with water	Binary (Yes/No)		[44]	
	Proportion of household habitat consisting of cropland	Continuous (Proportion)	Ecological mapping		
	GPS coordinates for each sample		Metadata		
Anthropogenic	Presence of livestock in household	Binary (Yes/No)	Household metadata		
	Total livestock in household	Continuous (Log[10])	Household metadata		
	Total humans in household	Continuous (Log[10])	Household metadata		
	Number of livestock in household	Continuous (Log[10])	Household metadata		Dropped in favour of individual livestock species
	Number of individual livestock species in household (cattle, goats, sheep, pigs, indigenous poultry, exotic poultry, lagomorphs, waterfowl, pigeons)	Continuous (Log[10])	Household metadata		
	Proportion of MDR in humans in household	Continuous (proportion)	AST data		Dropped in favour of median profile length
	Proportion of MDR in livestock in household	Continuous (proportion)	AST data		Dropped in favour of median profile length
	Proportion of MDR in household environmental samples	Continuous (proportion)	AST data		Dropped in favour of median profile length
	Median profile length in humans in household	Continuous	AST data		
	Median profile length in livestock in household	Continuous	AST data		
	Median profile length in household environmental samples	Continuous	AST data		
	Household garbage (rubbish) management	Categorical (Inside house, Outside house)	Household metadata		
	Household manure management	Categorical (Inside house, Outside house)	Household metadata		
	Position of household on wealth index	Continuous	Household metadata		Calculated based on methods used to create the DHS wealth index, but with a modification of the original variables included [22]
	Week of the year in which samples were collected	Continuous	Household metadata		
	Laboratory samples were tested in	Categorical (Kemri/UoN)	AST data		

Table 5.2. Ecological and anthropogenic covariates considered in this study.

		Prevalence				Wildlife Taxon				Feeding Niche				
		Wildlife	Human	Livestock	Environment	Avian	Rodent	Bat	Avian population	Fruit/Nectar	Invertebrate	Omnivore	Seed	Scavenger
Antimicrobial	**Amoxicillin C**													
Phenotype	Prevalence	6.42	9.04	5.35	10.94	8.14	3.24	0.00	15.21	0.00	7.76	2.60	8.65	13.22
	LCI (95% CI)	*4.35*	*6.13*	*3.69*	*7.39*	*5.32*	*1.11*	*NA*	*3.44*	*NA*	*3.06*	*0.65*	*4.95*	*1.79*
	UCI (95% CI)	*8.85*	*12.53*	*7.15*	*15.08*	*11.60*	*6.58*	*NA*	*33.48*	*NA*	*14.62*	*5.52*	*13.29*	*33.54*
	Ampicillin													
	Prevalence	33.80	46.45	33.87	35.95	34.39	31.06	35.62	45.02	25.97	31.73	33.11	38.03	53.32
	LCI (95% CI)	*29.72*	*41.44*	*30.29*	*30.22*	*29.05*	*24.11*	*19.75*	*25.02*	*11.33*	*21.71*	*25.94*	*31.02*	*28.69*
	UCI (95% CI)	*38.14*	*51.68*	*37.54*	*42.07*	*40.12*	*38.47*	*52.88*	*66.20*	*43.99*	*42.60*	*41.25*	*45.25*	*77.29*
	Chloramphenicol													
	Prevalence	5.74	7.80	4.83	5.10	8.15	0.64	6.94	9.89	10.91	7.63	1.91	8.00	13.49
	LCI (95% CI)	*3.88*	*5.12*	*3.30*	*2.73*	*5.32*	*0.02*	*0.95*	*1.34*	*2.29*	*2.99*	*0.40*	*4.35*	*1.63*
	UCI (95% CI)	*7.87*	*10.90*	*6.63*	*8.17*	*11.49*	*2.41*	*18.52*	*25.93*	*24.27*	*14.46*	*4.51*	*12.49*	*33.60*
	Ceftazidime													
	Prevalence	5.78	7.80	4.85	5.05	2.14	3.25	0.00	15.06	0.00	2.54	3.91	1.73	20.12
	LCI (95% CI)	*3.86*	*5.16*	*3.35*	*2.74*	*0.77*	*1.10*	*NA*	*3.32*	*NA*	*0.32*	*1.50*	*0.37*	*4.86*
	UCI (95% CI)	*7.99*	*10.93*	*6.65*	*8.04*	*4.13*	*6.57*	*NA*	*33.43*	*NA*	*7.01*	*7.56*	*4.08*	*43.00*
	Ciprofloxacin													
	Prevalence	4.09	3.13	3.61	3.86	4.30	3.27	7.14	5.05	3.71	5.12	5.22	3.47	6.73
	LCI (95% CI)	*2.51*	*1.50*	*2.31*	*1.86*	*2.28*	*1.17*	*0.94*	*0.16*	*0.12*	*1.45*	*2.32*	*1.28*	*0.23*
	UCI (95% CI)	*6.07*	*5.34*	*5.22*	*6.59*	*6.89*	*6.44*	*19.17*	*17.82*	*13.11*	*10.88*	*9.49*	*6.63*	*24.21*
	Gentamicin													
	Prevalence	5.78	6.55	8.77	11.72	7.10	1.95	14.13	5.03	7.39	8.82	3.27	7.49	6.72
	LCI (95% CI)	*3.85*	*4.18*	*6.69*	*8.08*	*4.47*	*0.40*	*4.17*	*0.11*	*0.93*	*3.58*	*1.14*	*4.05*	*0.23*
	UCI (95% CI)	*8.10*	*9.44*	*11.09*	*15.85*	*10.22*	*4.61*	*28.95*	*17.23*	*19.59*	*16.27*	*6.44*	*11.79*	*22.93*
	Cefotaxime													
	Prevalence	8.90	5.93	14.60	12.52	9.58	5.87	10.82	19.94	11.17	7.60	10.45	7.98	13.31
	LCI (95% CI)	*6.46*	*3.59*	*12.02*	*8.66*	*6.49*	*2.72*	*2.50*	*5.84*	*2.47*	*3.03*	*6.17*	*4.42*	*1.68*
	UCI (95% CI)	11.62	*8.72*	*17.46*	*16.92*	*13.33*	*9.97*	*24.24*	*38.95*	*24.58*	*14.53*	*15.66*	*12.33*	*34.91*
	Cefepime													
	Prevalence	8.82	5.88	14.57	12.44	3.25	1.85	0.00	5.18	0.00	5.24	2.63	2.29	6.47
	LCI (95% CI)	*6.42*	*3.66*	*11.92*	*8.77*	*1.54*	*0.37*	*NA*	*0.18*	*NA*	*1.54*	*0.74*	*0.69*	*0.17*
	UCI (95% CI)	*11.44*	*8.72*	*17.42*	*16.81*	*5.62*	*4.54*	*NA*	*17.85*	*NA*	*11.05*	*5.69*	*4.91*	*22.69*
	Nalidixic A													
	Prevalence	11.97	12.13	10.66	10.12	11.72	12.86	7.19	15.29	3.72	12.67	14.27	10.95	20.07
	LCI (95% CI)	*9.25*	*8.74*	*8.30*	*6.74*	*8.26*	*8.12*	*0.82*	*3.64*	*0.12*	*6.42*	*9.22*	*6.81*	*4.48*
	UCI (95% CI)	*14.99*	*15.80*	*13.15*	*14.09*	*15.60*	*18.42*	*19.01*	*33.92*	*12.69*	*20.60*	*20.54*	*16.01*	*43.16*
	Streptomycin													
	Prevalence	73.58	82.84	76.45	75.38	75.88	74.16	46.58	75.03	51.79	63.23	75.95	79.25	80.37
	LCI (95% CI)	*69.61*	*78.52*	*72.82*	*70.03*	*70.89*	*67.10*	*28.89*	*54.36*	*33.77*	*52.39*	*68.88*	*72.93*	*57.46*
	UCI (95% CI)	*77.45*	*86.63*	*79.73*	*80.45*	*80.74*	*80.67*	*65.24*	*90.65*	*69.71*	*73.22*	*82.42*	*84.88*	*95.46*
	Sulphonamide													
	Prevalence	78.36	84.74	74.45	75.86	79.07	78.15	82.17	65.11	85.32	73.34	78.49	79.99	93.54
	LCI (95% CI)	*74.61*	*80.89*	*71.07*	*70.57*	*74.38*	*71.45*	*66.22*	*44.45*	*69.85*	*63.31*	*72.02*	*73.77*	*77.13*
	UCI (95% CI)	*81.82*	*88.41*	*77.73*	*80.93*	*83.51*	*84.10*	*93.89*	*83.80*	*95.49*	*82.28*	*84.56*	*85.49*	*99.86*
	Tetracycline													
	Prevalence	33.65	45.17	45.42	44.04	33.59	32.31	21.43	60.34	7.38	31.61	29.23	37.35	66.52
	LCI (95% CI)	*29.53*	*39.92*	*41.48*	*38.08*	*28.28*	*25.31*	*8.74*	*37.90*	*1.01*	*21.97*	*22.24*	*30.37*	*42.15*
	UCI (95% CI)	*37.89*	*50.72*	*49.21*	*50.04*	*39.08*	*39.91*	*38.71*	*80.29*	*19.58*	*41.97*	*36.74*	*44.61*	*87.24*
	Trimethoprim													
	Prevalence	34.87	55.80	42.43	42.57	36.19	31.54	31.94	45.33	18.72	34.24	29.85	39.51	53.24
	LCI (95% CI)	*30.76*	*50.38*	*38.61*	*36.77*	*30.70*	*24.46*	*16.52*	*24.17*	*6.72*	*24.44*	*22.88*	*32.48*	*28.98*
	UCI (95% CI)	*39.25*	*61.38*	*46.34*	*48.81*	*41.88*	*39.18*	*49.53*	*67.26*	*35.31*	*45.04*	*37.46*	*46.74*	*76.57*

Table 5.3. Bayesian prevalence estimates (and 95% CI) for antimicrobial resistance to different drugs in wildlife (stratified by taxonomic group, and functional group), and across wildlife, livestock, human and environmental compartments. LCI and UCI are lower and upper confidence intervals respectively.

Model: MDR carriage in all isolates	Estimate	Std. Error	z value	P-value
Intercept	0.03	0.102	0.296	0.767
Environment	0.322	0.165	1.954	0.051
Human	0.662	0.155	4.273	<0.001
Livestock	0.284	0.128	2.222	<0.05
MEM1	0.148	0.067	2.225	<0.05
MEM2	-0.118	0.066	-1.781	0.075
MEM5	-0.13	0.063	-2.072	<0.05

Model: Profile length in all isolates	Estimate	Std. Error	z value	P-value
Intercept	1.095	0.031	35.8	<0.001
Environment	0.107	0.044	2.42	<0.05
Human	0.199	0.04	4.94	<0.001
Livestock	0.07	0.036	1.96	<0.05
MEM1	0.07	0.021	3.31	<0.001
MEM2	-0.042	0.02	-2.04	<0.05
MEM5	-0.049	0.019	-2.6	<0.01

Table 5.4. Estimated regression parameters, standard errors, z-values and P-values for the Binomial and Poisson generalised linear mixed effects models (GLMMs) presented in eqns (1) and (2), which compare the likelihood of wildlife, livestock, human and environmental compartments carrying multi-drug resistant (MDR)-*E. coli*, and the length of *E. coli* antibiotic profiles across wildlife, livestock, human and environmental compartments. MEM1, MEM2, and MEM5 indicate the spatial scales across which variation in MDR carriage or profile length occurs.

		Wildlife	Environmental	Human	Livestock
Diversity	**PR (profile richness)**				
Indices	Median	71	93	68	79
(Alpha)	LCI (95%)	*63*	*NA*	*62*	*70*
	UCI (95%)	*78*	*NA*	*73*	*88*
	SE (Shannon entropy)				
	Median	31.18	43.14	33.87	36.02
	LCI (95%)	*26.84*	*NA*	*30.7*	*30.36*
	UCI (95%)	*36.08*	*NA*	*37.06*	*42.5*
	SD (Simpson diversity)				
	Median	16.01	20.67	22.01	18.74
	LCI (95%)	*13.51*	*NA*	*17.99*	*15.89*
	UCI (95%)	*19.09*	*NA*	*22.31*	*22.55*
	BP (Berger-Parker)				
	Median	8.54	11.5	11.97	11.35
	LCI (95%)	*7.05*	*NA*	*10.56*	*9.61*
	UCI (95%)	*10.93*	*NA*	*13.62*	*13.29*

Table 5.5. Median values for each epidemiological compartment (wildlife, environmental, human, livestock) for four diversity indices (PR, SE, SD, BP) of phenotypic antimicrobial resistance. Environmental samples were resampled from, and as such 95% CIs were not calculated for this compartment. Wildlife had a statistically lower diversity across all levels of Alpha than all other compartments (P<0.001).

a)

		Wildlife	Environmental	Human	Livestock
Assymptotic	**Chao2**				
Richness	Estimate	272.54	349.89*	184.73	416.06
Estimator	LCI (95%)	244.63	304.99	204.86	378.1
	UCI (95%)	300.45	394.78	164.59	454.02
	ICE				
	Estimate	163.67	240.57	116.38	236.91
	LCI (95%)				
	UCI (95%)				
	Jacknife				
	Estimate	170.85	160.7	127.85	240.84
	LCI (95%)				
	UCI (95%)				

b)

Compartment	t	T	S_{obs}	S_{est}	Q_1	Q_2	q_0	g=1	g=0.85	g=0.8
Wildlife	440	440	105	273	66	13	0.15	8848	1572	1251
Environment	236	236	93	350	68	9	0.29	7464*	1410*	1154*
Humans	310	310	80	185	48	11	0.15	5000	896	702
Livestock	585	585	145	416	96	17	0.16	13979	2421	1947

Table 5.6. **a)** Asymptotic estimates for the total number of antimicrobial resistance profiles in each epidemiological compartment. Values in red indicate Chao 2 asymptotic estimates; *, estimator did not show asymptotic behaviour. **b)** Estimated sampling effort for each epidemiological compartment. Abbreviations are: t, number of samples collected; T, total number of incidences; S_{obs}, observed species richness; S_{est}, estimated asymptotic species richness, based on the Chao2 estimator; Q1, the number of species represented by exactly one sample (“uniques”); Q2, the number of species represented by exactly two samples (“duplicates”); q0, the probability that the next observed sample contains a species new to the survey (i.e., the proportion of species in the next sample that are new to the survey); g, target fraction of S_{est} that is to be reached. The entries in each “g” column represent the number of additional samples required to detect that proportion of S_{est}. The estimators for Environment did not show asymptotic behaviour, and as such values marked with * are interpreted with caution.

Model: MDR carriage in wildlife	Estimate	Std. Error	z value	P-value
Intercept	-1.6578	0.7756	-2.137	<0.05
Avian_VertFishScav	2.3775	0.9961	2.387	<0.05
Avian_Invertebrate	1.3414	0.8217	1.632	0.102
Avian_Omnivore	2.0119	0.8494	2.369	<0.05
Avian_Seed	1.842	0.7926	2.324	<0.05
Bat_Fruit	2.3151	0.976	2.372	<0.05
Bat_Invertebrate	0.978	0.9477	1.032	0.3021
Rodent_Omnivore	1.524	0.7997	1.906	0.0567
MEM1	-0.2877	0.1212	-2.373	<0.05

Model: AMR-profile length in wildlife	Estimate	Std. Error	z value	P-value
Intercept	0.65937	0.20319	3.245	<0.01
Avian_VertFishScav	0.58595	0.24446	2.397	<0.05
Avian_Invertebrate	0.34555	0.21684	1.594	0.111
Avian_Omnivore	0.57958	0.22161	2.615	<0.01
Avian_Seed	0.43354	0.20869	2.077	<0.05
Bat_Fruit	0.45743	0.26255	1.742	0.081
Bat_Invertebrate	0.20251	0.25787	0.785	0.432
Rodent_Omnivore	0.26578	0.21093	1.26	0.208
MEM1	-0.11512	0.03607	-3.192	<0.01
MEM8	-0.11512	0.03186	3.255	<0.01
MEM10	0.10369	0.03219	1.985	<0.05
MEM18	-0.04822	0.03149	-1.531	0.126
MEM25	-0.05295	0.02912	-1.819	0.069
MEM27	-0.07839	0.02889	-2.713	<0.01
Laboratory (UoN)	0.19678	0.07282	2.702	<0.01

Table 5.7. Estimated regression parameters, standard errors, z-values and P-values for the Binomial and Poisson generalised linear mixed effects models (GLMMs) presented in eqns (3) and (4), which compare the likelihood of different wildlife functional groups carrying multi-drug resistant (MDR)-*E. coli*, and the length of *E. coli* antibiotic profiles across wildlife functional groups. MEMs indicate the spatial scales across which variation in MDR carriage or profile length occurs. Increasing MEM numbering indicates increasingly finer spatial scales.

Model: MDR carriage in seed-eating birds	Estimate	Std. Error	z value	P-value
Intercept	-5.4935	2.3398	-2.348	<0.05
Total Cattle	3.4136	1.0158	3.361	<0.001
Human ABG	1.2222	0.5443	2.245	<0.05
Livestock ABG	0.1056	0.2893	0.365	0.715
Manure (outside house)	2.5294	1.4222	1.779	0.075
Garbage (outside house)	4.7585	2.0421	2.32	0.02
Garbage (outside house): Human ABG	-1.0513	0.5332	-1.972	<0.05
Manure (outside house): Livestock ABG	-0.9655	0.4708	-2.051	<0.05

Model: MDR carriage in rodents	Estimate	Std. Error	z value	P-value
Intercept	-4.3039	1.7504	-2.459	<0.05
Human ABG	1.3059	0.5383	2.426	<0.05
Livestock ABG	0.4085	0.1942	2.104	<0.05
Manure (outside house)	2.9078	1.265	2.299	<0.05
Garbage (outside house)	1.4198	1.6627	0.854	0.393
Laboratory (UoN)	-2.0261	1.1738	-1.726	0.084
Garbage (outside house): Human ABG	-1.0043	0.5821	-1.725	0.085
Manure (outside house): Human ABG	-0.5909	0.329	-1.796	0.073

Table 5.8. Estimated regression parameters, standard errors, z-values and P-values for the Binomial generalised linear mixed effects models (GLMMs) presented in eqns (5) and (6), which compare risk factors for the likelihood of multi-drug resistant (MDR)-*E. coli* carriage in seed-eating birds and rodents.

5.9 Supplementary Figures

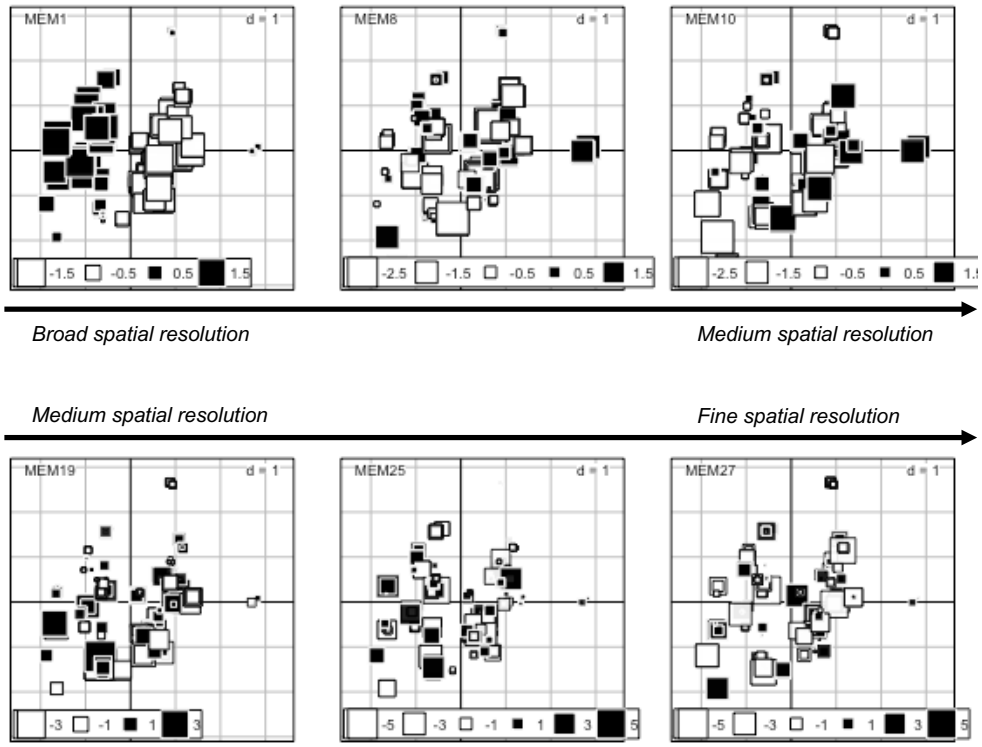


Figure 5.1. Plot showing distance-based moran's eigenvector maps (dbMEM) eigenvectors modelling significant spatial variation at different scales across Nairobi. Increasing numeric magnitude of MEMs represents a gradient of broad to fine-scale spatial structure. As such, in this example MEM1, MEM8 and MEM10 model broad through to medium scale spatial resolution, and MEM19, MEM25 and MEM27 model medium through to fine-scale spatial resolution across Nairobi. Black and white blocks represent the GPS location of wildlife samples.

$$\mathrm{MDR}_{ij} \sim \mathrm{Binomial}(\mu_{ij})$$

$$\mathrm{E}(\mathrm{MDR}_{ij}) = \mu_{ij}$$

$$\frac{\log(\mu_{ij})}{1 + \log(\mu_{ij})} = \mathrm{TaxonFunc}_{ij} + \mathrm{HomeRange}_{ij} + \mathrm{TaxonFunc}_{ij} \times \mathrm{HomeRange}_{ij} + \mathrm{MEM1}_{ij} + \mathrm{Laboratory}_{ij} + \mathrm{Sublocation}_{i}$$

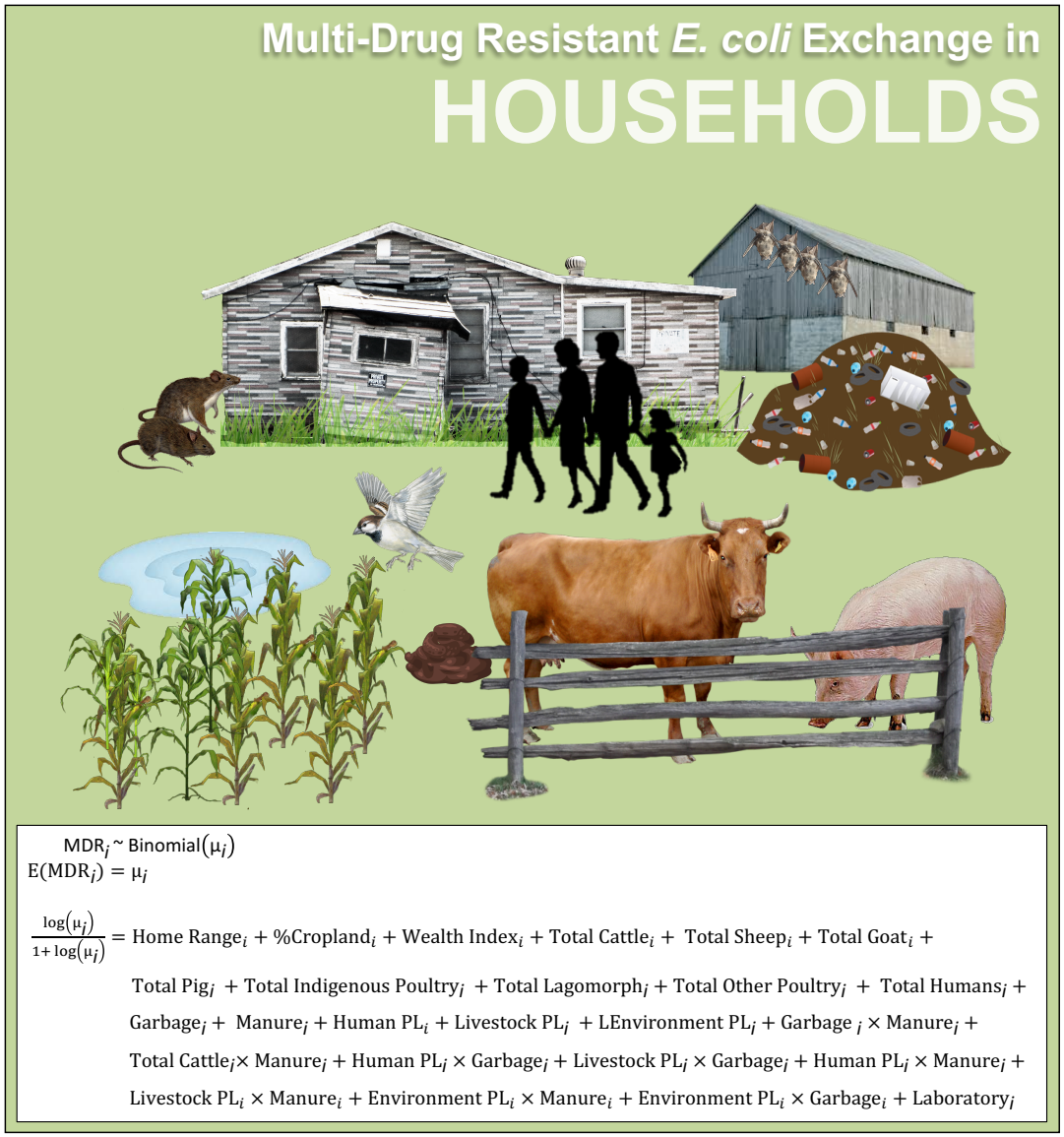
$$\mathrm{Profile\ Length}_{ijk} \sim \mathrm{Poisson}(\mu_{ijk})$$

$$\mathrm{E}(\mathrm{Profile\ Length}_{ijk}) = \mu_{ijk}$$

$$\log(\mu_{ijk}) = \mathrm{Taxon}_{ijk} + \mathrm{HomeRange}_{ij} + \mathrm{MEM1}_{ijk} + \mathrm{MEM8}_{ijk} + \mathrm{MEM10}_{ijk} + \mathrm{MEM19}_{ijk} + \mathrm{MEM25}_{ijk} + \mathrm{MEM27}_{ijk} + \mathrm{TaxonFunc}_{ij} \times \mathrm{HomeRange}_{ij} + \mathrm{Laboratory}_{ijk} + \mathrm{Household}_{i} + \mathrm{Sublocation}_{j}$$

$$\mathrm{Household}_{i} \sim N(0, \sigma^2)$$

Figure 5.2. Mathematical representation of the full Binomial and Poisson generalised linear mixed effects models (GLMMs, before selection) for equations (3) and (4). These models explain variation in seed-eating bird and a rodent MDR-*E. coli* carriage, as a function of household-level ecological and anthropogenic covariates.



$$\mathrm{MDR}_i \sim \mathrm{Binomial}(\mu_i)$$
$$\mathrm{E}(\mathrm{MDR}_i) = \mu_i$$

$$\frac{\log(\mu_i)}{1+\log(\mu_i)} = \text{Home Range}_i + \%\text{Cropland}_i + \text{Wealth Index}_i + \text{Total Cattle}_i + \text{Total Sheep}_i + \text{Total Goat}_i + \text{Total Pig}_i + \text{Total Indigenous Poultry}_i + \text{Total Lagomorph}_i + \text{Total Other Poultry}_i + \text{Total Humans}_i + \text{Garbage}_i + \text{Manure}_i + \text{Human PL}_i + \text{Livestock PL}_i + \text{LEnvironment PL}_i + \text{Garbage}_i \times \text{Manure}_i + \text{Total Cattle}_i \times \text{Manure}_i + \text{Human PL}_i \times \text{Garbage}_i + \text{Livestock PL}_i \times \text{Garbage}_i + \text{Human PL}_i \times \text{Manure}_i + \text{Livestock PL}_i \times \text{Manure}_i + \text{Environment PL}_i \times \text{Manure}_i + \text{Environment PL}_i \times \text{Garbage}_i + \text{Laboratory}_i$$

Figure 5.3. Diagrammatic and mathematical representation of the full Binomial generalised linear mixed effects model (GLMM, before selection) for equations (5) and (6). These models explain variation in seed-eating bird and a rodent MDR-*E. coli* carriage, as a function of household-level ecological and anthropogenic covariates.

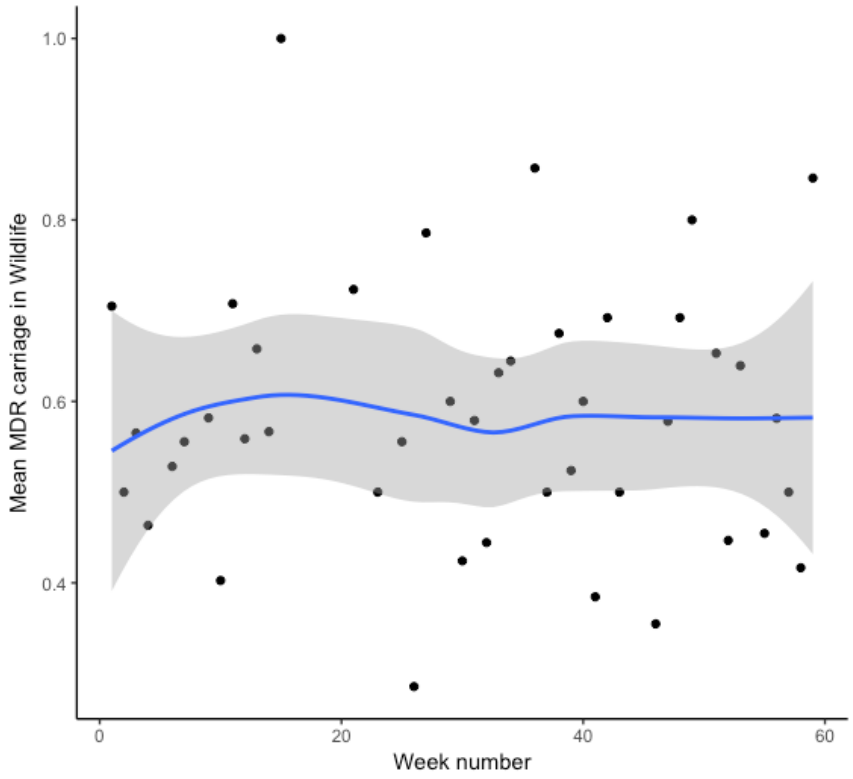


Figure 5.4. Plot showing variation in mean multidrug resistant (MDR)-*E. coli* carriage in wildlife sampled per week, according to the week of the study in which sampling was conducted (blue line: mean, grey shaded area: 95% CI's around the mean).

Chapter 6

Changes in Urban Land Use Drive the Structure of Bacterial Genetic Communities in Wildlife Hosts

Data used in the analyses relating to mobile genetic elements (MGEs) contained within this chapter is available via an open access repository held by the University of Liverpool (http://datacat.liverpool.ac.uk/id/eprint/526).

6 Deterministic Processes Structure Bacterial Genetic Communities across an Urban Landscape

6.1 Abstract

Land-use change is predicted to act as a key driver of zoonotic disease emergence through human exposure to novel microbial diversity, but evidence for the effects of environmental change on microbial communities in vertebrate hosts is lacking. Using molecular data sourced from the core and accessory components of *E. coli* whole genome sequences collected from wildlife, livestock and humans in households across Nairobi, I investigate how bacterial populations nested within host communities at interfaces respond to urban land-use change, and how the population structure of *E. coli* varies between hosts. Initially, the response of freely exchangeable mobile genetic elements (MGEs, sourced from different wildlife hosts) to perturbations in host community structure and other forms of anthropogenic variation within household interfaces is investigated. By modelling the diversity of two sets of bacterial genes (those encoding virulence and antimicrobial resistance (AMR)) against ecological and anthropogenic forms of urban environmental change, I demonstrate that communities of bacterial genes in avian hosts are shaped by the structure of co-existing host communities, and the habitat within which they exist. I then use core genome multilocus sequence typing (cgMLST) to characterise the population structure of *E. coli* in urban wildlife, demonstrating that most of the explained genetic variation between hosts arises through differences in their feeding ecology. In showing that deterministic (i.e. non-random) processes are important in structuring bacterial genetic communities in urban wildlife, I demonstrate that it is possible to link epidemiological processes to ecological and anthropogenic drivers across an urban landscape. These findings suggest that it should be possible to forecast the effects of urban land-use change on microbial diversity.

6.2 Introduction

As a consequence of anthropogenic activities and the fragmentation of natural habitats, rapid, unplanned urban development in cities such as Nairobi results in high heterogeneity of land use. The complex mosaic of remaining habitats exist along a gradient of ecological complexity, and as such offer a variety of natural and artificial niches that shape the composition of wildlife host communities present. Where wildlife co-exist with livestock and humans, the host structure of these interfaces are determined by a combination of habitat and anthropogenic determinants (such as socioeconomics). As postulated in Chapter 4, such profound effects on the structure of host communities would be expected to produce perturbations within their in-situ microbial communities. However, there is little empirical evidence that directly links changes in the function of abiotic and biotic systems to the structure of host communities, and dynamics of microbes living within them. In this chapter, *E. coli* genetic data collected from vertebrate hosts in Nairobi, is used to detect and describe perturbations in microbial communities within individuals in the wildlife population across the city.

Recent advances in sequencing technology, such as whole genome sequencing (WGS), provide accurate classification of bacterial strains based on stable housekeeping genes that constitute the core genome, and offer the potential to study the community of genes carried on mobile genetic elements (MGEs) within prokaryote genomes. MGE-borne genes can be horizontally transferred between organisms via intercellular movements, and may confer adaptive functional traits such as antimicrobial resistance and virulence [202]. Bacteria can exchange MGEs through three forms of intercellular movement; transformation (direct chromosomal transfer between closely related bacteria), transduction (via bacterial viruses called bacteriophages) and conjugation (on plasmids or transposons) [202]. The distribution of MGE-borne genes amongst bacteria can therefore provide insight into the evolutionary history and community structure of these micro-organisms, an approach that has been successfully used in conjunction with typing tools and time-scaled evolutionary analyses to infer bacterial transmission between hosts [195,196,203]. The wealth of genetic data generated by WGS could therefore provide an optimal approach to identify key drivers (such as land-use change) that influence the structure of

bacterial populations at wildlife-livestock-human interfaces, and assist in untangling the complexity of epidemiological processes, regardless of the taxonomic distance between hosts.

As a ubiquitous, genetically diverse bacterium, *E. coli* is frequently used as an exemplar organism to investigate epidemiological connectivity between vertebrate host populations. Both commensal and pathogenic strains of *E. coli* are transferred in a similar way to many gastrointestinal pathogens (i.e. through the faecal-oral route) and thus represent a widely-used and ideal surrogate system for investigating transmission [80,204–206]. Although certain pathogenic strains (such as *E. coli* O157) represent emerging infectious diseases of public health importance, it has been suggested that commensal strains are more appropriate for investigating transmission pathways, since they are ubiquitous and host resistance is less likely to have evolved [207].

By utilising *E. coli* whole genome sequences, this chapter aims to address two questions. Initially, building upon results from Chapter 4 in which gradients of ecological and anthropogenic change were identified across Nairobi, the response of microbial communities in wildlife hosts to urban land-use change is investigated; specifically, whether determinism in community structure of bacterial MGEs is driven by changes in host community structure and anthropogenic factors. MGE diversity is utilised as a proxy for parasite diversity, with the view that the ability of MGEs to move relatively freely between bacterial cells through horizontal gene transfer mimics, to an extent, the movement of directly-transmitted parasites between hosts. Two classes of MGE were selected (those encoding antimicrobial resistance (AMR) and virulence traits), each of which would be expected to respond differently to urban change. Contamination of the external environment with antimicrobial resistant bacteria excreted from humans and livestock treated with antimicrobials (e.g. through sewage effluent or faeces), is considered an important route of wildlife exposure to AMR[121]. As such, if *E. coli* in wildlife are under higher selective pressure to adopt genes encoding AMR in urban areas where greater volumes of antibiotics are consumed, and antibiotic use is more widespread[208,209], the community structure of MGEs encoding AMR would be hypothesised to respond to anthropogenic changes, rather than ecological variability in host communities. In contrast, the diversity of genes encoding virulence traits (for which *E.* coli in wildlife are assumed not to be subjected to such strong anthropogenic selection pressure) would be hypothesised to

reflect changes in avian host community structure – following the broadly accepted principal that host and microbial community diversity are correlated [156,157], as avian host species diversity increases, their pool of *E. coli*-borne virulence-associated MGEs should become more diverse. Hence, aside from investigating processes underlying determinism in bacterial genetic diversity, studying the diversity of two sets of genes that confer selective pressure to bacteria will enable me to assess whether an association exists between urban land use and the genetic determinants of bacterial selection, with potential implications for human and animal health [202].

Secondly, building upon the results of Chapter 5 where signals of epidemiological connectivity were identified between synanthropic wildlife, livestock and humans at household interfaces, the epidemiological structure of wildlife-borne *E. coli* is investigated at a city-wide scale and related to the population structure of *E. coli* in other epidemiological compartments (livestock, humans and the outside environment). Specifically, multivariate methods are applied to core-genome multi-locus sequence typing (cgMLST) of *E. coli* to test whether host/source identity and/or geographic distance ("isolation by distance") drive differentiation of *E. coli* populations across the city, and how this varies between epidemiological compartments.

6.3 Methods

6.3.1 Data Collection

Sampling

Faecal samples (n=2081) were collected from 78 wildlife species (birds and mammals, n=794), 13 livestock species (n=677), humans (n=333) and the outside environment (n=277) from 99 households across Nairobi, that were participating in this project (Chapter 3, Suppl. Table 3.1). Study design (stratification and selection of households) is explained in detail in Chapter 2.4. Wildlife and livestock samples were obtained from a single extra house, making the total number of households for which either wildlife or livestock were considered to be 100.

Up to 20 rectal swabs were obtained from livestock species present in the household compound (ensuring that all species present were represented in the household sample). Wildlife samples were obtained by a range of taxon-specific trapping methods, details of which are included in Chapter 3.5. All wildlife, livestock and human sampling was carried out by trained veterinarians or clinicians. Environmental samples were collected from livestock pens, soil, puddles and flowing water around the household compound, as described in Chapter 3.5.

Laboratory procedures

Bacterial culture and isolation of *E. coli* from samples is described in Chapter 3.6. DNA was extracted from bacterial isolates using commercial kits (Purelink® Genomic DNA Mini Kit, Invitrogen, Life Technologies, Carlsbad, California) and transported under licence to The Wellcome Trust Centre for Human Genetics, Oxford, UK.

6.3.2 Data Processing

Whole genome sequencing (WGS) was carried out at the Wellcome Trust Centre for Human Genetics on the Illumina HiSeq 2500 platform. 150 base-pair paired-end reads were generated and standard quality control checks were performed. Short-read WGS data were processed using the semi-automatic pipeline developed by the Modernising Medical Microbiology Oxford (MMM) group. Reads were first trimmed to remove remnant adaptor sequences. Speciation was performed using Kraken [210] with an in-house database of bacterial reads downloaded from the NCBI sequence read archive (www.ncbi.nlm.nih.gov/sra/). This was followed by human and virus read removal. A reference sequence for mapping short-reads was automatically selected based on the speciation results (AE014075.1 for *E.coli*). Samples deemed as non-*E. coli* were excluded from further analysis. *De novo* assembly of *E. coli* isolates was performed using SPAdes v3.6 [211]. Potentially mixed *E. coli* samples were identified as those with an unusually large assembly size (greater than 6 megabases (Mb), compared to a reported *E. coli* genome size range of 4.56-5.93Mb) [212] and were removed from the dataset. The assemblies were run through the batch upload

mode of the Centre for Genetic Epidemiology web interface hosted by the Technical University of Denmark (https://cge.cbs.dtu.dk/services/cge/) which performs speciation analysis [213], multilocus sequence typing (MLST) [214], detection of resistance genes [215], detection of virulence genes [216] and detection and typing of plasmids [217].

Typing

Core genome multilocus sequence typing (cgMLST) was used to analyse genomic relatedness utilising pairwise comparisons of genetic loci within the core genome. cgMLST relies on the construction of a species-specific scheme, where target genes across the entire core genome form the basis of genetic comparison of alleles [218]. A stable cgMLST scheme for *E. coli* was created using the software SeqSphere+ (version 2.3; Ridom GmbH, Münster, Germany). 370 complete *E. coli* genomes from NCBI GenBank were used to define a set of 1092 target genes, and strain CFT073 (GenBank accession no. NC_004431.1) was selected as the reference genome. The set of target genes was determined using the target definer function in SeqSphere+, and default filters were used for the inclusion of genes in the target scheme. The following filters were applied to the reference genome; all gene sequences < 50 base pairs were discarded, genes without a start codon were removed, all gene sequences without a stop codon, with more than one codon or a stop codon not at the end of the gene sequence were discarded, all gene sequences with fragments that occurred in multiple copies in a genome (with identity ≥ 90% and more than 100 bases overlap) were discarded, and if two gene sequences overlapped by more than 4 base pairs, the shorter gene sequence was discarded. The remaining gene sequences were compared against 370 query genomes using BLAST, with the final cgMLST scheme (comprising 1092 gene sequences) being composed of gene sequences bearing a sequence identity of ≥ 90% and 100% overlap between reference and query genomes (as used in previous studies [219–221]). Gene sequences of the reference genome that were common in all query genomes with a sequence identity of ≥90% and 100% overlap formed the final cgMLST scheme (suppl. Table 1). WGS *E. coli* assemblies from wildlife, livestock, humans and the outside environment were blasted against this scheme, generating 1485 cgMLST profiles.

6.3.3 Data Analysis: Variables

For the purposes of analyses, the sources from which isolates were derived were split into host/source groups. Wildlife isolates were grouped by the feeding-ecology related functional groups considered throughout this thesis (see Chapter 3.5, Table 3.2). However, due to limited sample size, primates and carnivores were excluded from analyses conducted in this chapter. Livestock isolates were grouped into the following broad, taxonomic groups, to account for small sample sizes amongst certain species: bovid (cattle), small ruminant (sheep and goats), pigs, chickens (indigenous and exotic), and fowl (turkeys, guinea-fowl, ducks and geese). Environmental isolates were grouped into two categories: outside environment (soil and water samples collected outside the house but within the compound), and livestock pens (samples collected using boot-socks from livestock-keeping areas). All analyses were conducted in R [136].

6.3.4 Data Analysis: Responses of microbial genetic communities to urban change

To test whether microbial genetic communities in wildlife are deterministically structured according to their response to urban land-use change, two sets of mobile genetic elements (MGEs) that would be expected to respond to different drivers of urban change were chosen: virulence genes, and AMR genetic determinants. α-diversity (counts of each set of MGEs, thus representing 'richness' of virulence or AMR genes) was calculated for each individual host, and compared against environmental and anthropogenic determinants of urban land-use change. Because sampling and community-characterisation of avian hosts in the field study was more complete than for any other taxonomic group of wildlife, only avian-derived isolates were considered in these analyses. The following set of household-level environmental and anthropogenic determinants (common to those used in previous chapters, and described in detail in Chapter 3) were selected as indicators of urban land-use change: *host functional group membership*, *α-diversity of avian species (richness), habitat (biotic) diversity*, *artificial land-use cover*, *wealth and ruralness indices*, *livestock keeping*, and *human density*. Separate models, in which the effects of host spatial structure were tested against genetic diversity, were also developed. As in previous

chapters, distance-based Moran's eigenvector maps (dbMEMs) were used to represent spectral decomposition of the spatial relationship between isolates across multiple scales (see Chapter 3.7.4 for a full explanation of this approach). Of the resulting dbMEM eigenvectors, only those modelling significant spatial variation in genetic dissimilarity of *E. coli* (determined by backwards stepwise selection performed on a linear model of all eigenvectors modelling positive spatial variation between isolates) were included in final models. α-diversity of virulence and AMR genes were regressed against variables in generalised linear mixed effects models (GLMMs), fitted with Poisson distributions in the R package 'lme4' [107]. For each set of response and explanatory variables, data exploration was carried out following the protocol described in Zuur *et al.* [98]. To account for the dependency structure of the data, the household and/or sublocation in which samples were collected were included as random effects (Chapter 3.7.3, Figure 3.3). Final models were constructed using stepwise, backwards elimination from the full model. Model assumptions were verified by plotting residuals versus fitted values, and by assessing models for overdispersion. Non-linear relationships were checked by fitting a generalized additive model (GAM) between the response and explanatory variables, featuring a nonlinear smoother, in R package 'mgcv' [109]. The residuals were also assessed for spatial dependency by plotting them against geographic coordinates, and examining the results of a semivariogram. Where possible, marginal regression coefficients of multiple determination were reported (marginal R^2).

Preliminary data exploration indicated substantial zero-inflation in the response variable α-diversity of AMR genes (i.e. many samples where no AMR genes were detected), and as such a zero-inflated Poisson model (ZIP) was initially fitted to the data (56% of data comprising the response variable were zeros). However, residuals from the optimal ZIP model obtained through step-wise selection showed considerable overdispersion (dispersion statistic: 3). Dispersion parameters were stabilised by fitting a zero-inflated negative-binomial "type 1" generalised linear model (GLM) in R package 'glmmADMB' [108]. Rather than following a standard negative binomial fit, this family of models are parametrised in the same way as, and with an equivalent mean-variance relationship to a quasi-Poisson model.

6.3.5 Data Analysis: Genetic structure of *E. coli* populations in wildlife, and in relation to livestock and humans across Nairobi – isolation by environment and isolation by distance

In this section, multivariate approaches were used to explore the genetic structure of *E. coli* in wildlife across Nairobi, as determined by host/source and geographic distance. Exploratory analysis of the cgMLST dataset was carried out using multivariate clustering methods, which represent a viable approach to examine the structure of genetic populations, offering an alternative to Bayesian clustering techniques (such as STRUCTURE) which may not be appropriate for use with clonal organisms such as *E. coli*. Discriminant Analysis of Principal Components (DAPC) is a multivariate statistical method that partitions sample variance into between- group and within- group components, to maximize discrimination between groups [222]. Discriminant functions, which are generated from a principal component analysis (PCA) performed on the original data (alleles) and subsequent discriminant analysis (DA), enable samples to be assigned membership of genetic groups. DAPC is typically used to define genetic groups without prior knowledge, where a sequential K-means clustering algorithm is used to find clusters, which are compared using the Bayesian Information Criterion (BIC) to find an optimal number of genetic groups that describe the data. This approach was used to infer population structure of *E. coli* cgMLST results from all wildlife samples without prior knowledge. However, DAPC is also capable of identifying genetic clustering between biologically predefined groups, and as such, *a-priori* taxonomic groupings were used to infer the population structure of *E. coli* from cgMLST results, as implemented at two levels; i) between wildlife functional groups, and ii) between wildlife functional groups, humans, select livestock groups and the outside environment. All DAPC analyses were implemented in the R package 'Adegenet' [223].

DAPC is very sensitive to the selection of principal components (PCs), which can have a substantial impact on the results of the analysis. Selecting too few PCs can result in unexplained variance, whilst including too many can lead to overfitting of the discriminant functions, and erroneous assignment of clusters [224]. As such, two optimization procedures available in the Adegenet, were used to evaluate and select an optimal number of PCs to retain [224]. The α-score represents the difference between the proportion of successful reassignments of the analysis (observed

discrimination) and values obtained using random groups (random discrimination), and is computed by re-running the DAPC analysis with randomised groups. Cross-validation represents an alternative approach in which the data is divided into training and validation sets, which contain 90% and 10% of the data respectively. The accuracy with which DAPC performed on the training set, retaining variable numbers of PCs, can predict the group membership of samples in the validation set, is used to identify an optimal number of PCs to retain.

The relationship between *E. coli* population structure, host/source groups and spatial variation was further explored using Analysis of Molecular Variance (AMOVA), pairwise estimates of Weir and Cockerham's F_{ST}, and constrained ordination. AMOVA, implemented in the R package 'Poppr' [225], was used to partition the genetic variance of *E. coli* within and between wildlife functional groups and households, thus using our study design to test whether taxonomic functional groups act as barriers to *E. coli* exchange within household interfaces, and whether *E. coli* is exchanged between households. Weir and Cockerham's F_{ST} represents a measure of population substructure, which, unlike other fixation indices is unbiased to variation in sample size, and is therefore less likely to generate biased estimates [226,227]. Pairwise F_{ST} estimates were generated in the R package 'hierfstat' [228], and used to explore genetic differentiation between epidemiological compartments.

To test the hypotheses that genetic differentiation between *E. coli* isolates is driven by source identity and/or geographic distance between sources, and to explore differences in the spatial distribution of *E. coli* by source groups, distance-based redundancy analysis (db-RDA, a form of constrained canonical analysis performed on distance matrices – further details are included in Chapter 3.7.2) was used to develop and test a series of models. The first set of models compared the effect of measured and unmeasured spatial variation on genetic dissimilarity between isolates, with the inclusion of isolates being dependent upon the epidemiological compartments between which models were compared:

i) Isolates originating from wildlife (avian, rodent and bat functional groups),

ii) isolates originating from livestock [fowl (ducks, geese, turkeys, guinea-fowl), chickens, small ruminants (sheep and goats), bovids and pigs],

iii) isolates originating from humans,

iv) isolates originating from the outside environment (soil, and livestock pens).

v) isolates originating from all compartments combined (wildlife, livestock, humans and the outside environment).

The response variable consisted of a binary matrix representing the presence/absence of all cgMLST genes for each isolate within the compartment, which was transformed into a pairwise genetic distance matrix, with dissimilarity between isolates represented by the Jaccard index. dbMEMs were used as spatial explanatory variables, as previously described, with each epidemiological compartment (and therefore model) having its own unique set of dbMEMs. Of the resulting eigenvectors, only those modelling significant spatial variation in genetic dissimilarity of *E. coli* (determined by forward selection performed on a canonical model of all eigenvectors modelling positive spatial variation between isolates in that group) were included in the global RDA model for each compartment. To avoid correlation between host identity and spatial scales, the effect of host/source identity was removed by fitting it as a conditional term in the model [145]. Providing the global model was statistically significant, forward selection with a double stop criterion was used to determine an optimal canonical model for each compartment [145].

The second set of models were aimed at determining the influence of host/source identity on genetic dissimilarity of *E. coli* once spatial variation and household/sublocation hierarchical structure had been removed as potential confounders:

i) isolates originating from wildlife (as above),

ii) isolates originating from livestock (as above),

iii) isolates originating from all compartments combined (as above).

The response variable for these models consisted of a genetic dissimilarity matrix (Jaccard Index) and the explanatory variable host/source identity. To account for the confounding effects of spatial variance and hierarchical structure derived from the experimental design, the variance attributed to significant dbMEM eigenvectors

identified in the first set of models, and the covariables 'household' and 'sublocation' were removed in each canonical model by fitting them as conditional terms. For all canonical models, adjusted regression coefficients of multiple determination were reported (R^2_{adj}). db-RDA models were executed in the R package 'vegan', using function 'capscale' [104].

The use of multivariate analyses based upon a matrix of pairwise distances (such as the db-RDA approach described above) has been criticised by Warton *et al.* [229] for misspecifying the mean-variance relationship between response taxa (in this case *E. coli* genes). As such, a multivariate generalised linear modelling (GLM) approach developed by Warton *et al.* [229] was attempted, utilising the 'manyglm' function within R's 'mvabund' package [229,230]. This method involves fitting separate, univariate GLMs, which in this case, relate the presence of a single gene in each isolate to the host/source or spatial explanatory variables. Unfortunately, this approach proved too computationally intensive to perform with the large number of genes present in all cgMLST schemes, and as such was not implemented for this analysis.

6.4 Results

6.4.1 Responses of microbial genetic communities to gradients of urban change

Genes encoding virulence (n=90) and AMR (n=91) were obtained from a total of 413 wildlife-borne *E. coli* whole genome sequences. Of these, 258 isolates were from birds, 21 from bats, 3 from carnivores, 4 from primates and 157 from rodents. Only avian-derived sequences from individual birds trapped within household compounds (n=239) were considered for inclusion in models for virulence and AMR, and sequences originating from pooled avian populations were excluded from the analysis. The α-diversity of virulence genes in birds was modelled against species (*host functional group*) and household-level (*avian species richness*, *habitat diversity*, *artificial land-use cover*, *livestock-keeping, human density, wealth and ruralness*) explanatory variables in a Poisson GLMM. Intercept-only mixed-effects models

indicated that household and sublocation explained very little of the variance in diversity of virulence genes (5.5% and 7.8% respectively). Model fit was improved by excluding sublocation, and as such only household dependency was included as a random effect in the full model. Backwards step-wise selection resulted in an interaction term being fitted between the fixed covariates *avian species richness* (continuous) and *host functional group* (categorical with four levels), with *habitat diversity* (continuous) and *livestock-keeping* (categorical) as separate fixed effects, in a model of the form:

$$\text{Virulence gene diversity}_{ij} = \text{Avian Species Richness: Functional Grp}_{ij} + \text{Habitat diversity}_{ij} + \text{Livestock.keep}_{ij} + \text{Household}_i \quad \text{(eqn 1)}$$

where virulence gene diversity$_{ij}$ is the jth observation in household i, and i = 1,...,100, and household$_i$ is the random intercept, which is assumed to be normally distributed with mean 0 and variance σ^2. Model validation did not show substantial departure from the assumptions of the model, and all fixed terms except livestock-keeping were statistically significant, showing that α-diversity of virulence genes varies between

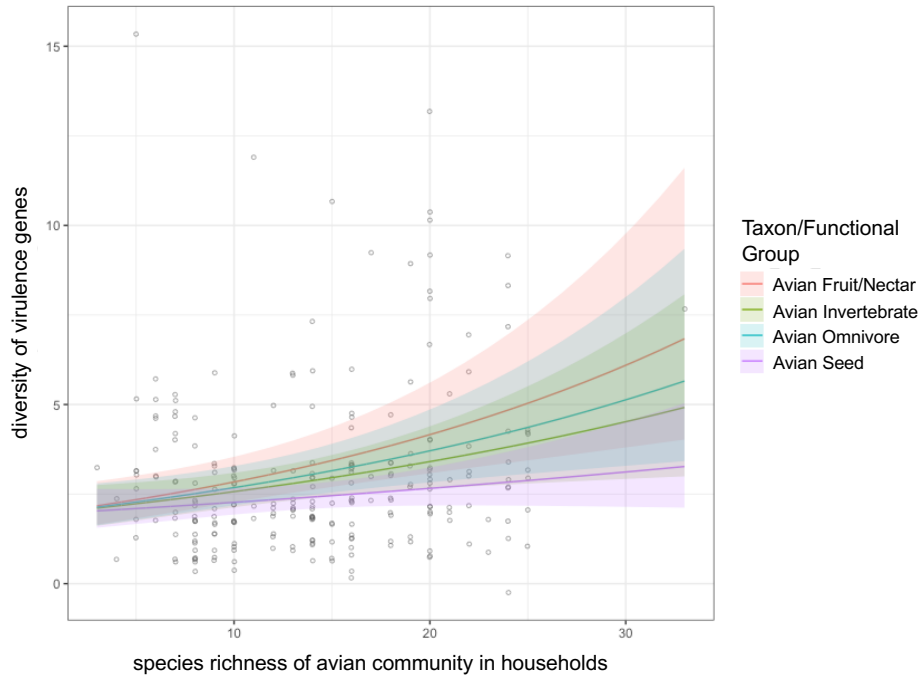


Figure 6.1. Fit of Poisson generalized linear mixed effects model (GLMM) in eqn (1), relating diversity of virulence genes in avian-borne *E. coli* to determinants across households. This graph shows how α-diversity (richness) of virulence genes in avian hosts increases as a function of interface host-community richness and functional group membership. Shading around lines represents 95% confidence intervals.

host functional group identity, and increases with α-diversity of household avian communities (Table 6.1, Figure 6.1). This relationship was present for all functional groups except seed-eating birds. Habitat diversity showed a weakly significant inverse relationship with virulence gene diversity. In total, the optimal model explained a relatively low proportion of variation in virulence gene diversity (marginal R^2: 0.1).

To further explore determinants of virulence gene diversity in seed-eating birds (which, as synanthropic species, constituted the largest and most well-distributed avian functional group), two Poisson GLMMs were built considering only the genetic diversity of sequences derived from this functional group. This also had the effect of removing variation associated with functional group membership. The first model used the same set of fixed and random covariates as the full avian model, whilst the second considered a set of spatial explanatory variables derived from dbMEMs. Step-wise selection for the first model resulted in retention of a single fixed covariate *habitat diversity* (continuous):

$$\text{Virulence gene diversity}_{ij} = \text{Habitat diversity}_{ij} + \text{Household}_i$$

(eqn 2)

Habitat diversity had a statistically significant inverse relationship with diversity of virulence genes in seed-eating birds, but explained little of the variance in virulence gene diversity (marginal R^2: 0.05). The spatial model was constructed from 17 dbMEM eigenvectors as fixed covariates, which accounted for broad (MEM1), medium and fine (MEM17) scale spatial structure among seed-eating birds across the city, and included household as a random effect. Backwards step-wise selection resulted in a model of the following form:

$$\text{Virulence gene diversity}_{ij} = \text{MEM2}_{ij} + \text{MEM6}_{ij} + \text{MEM10}_{ij} + \text{MEM13}_{ij} + \text{MEM16}_{ij} + \text{Household}_i$$

(eqn 3)

The results of this model demonstrate that the genetic diversity of virulence in a single functional group of birds varies across broad (MEM2-MEM6), moderate (MEM10) and fine (MEM13-MEM16) spatial scales in the city (Table 6.1, suppl. Figure 6.1).

This model explained a relatively large proportion of variance in virulence gene diversity across the city (marginal R^2: 0.27).

Model terms	Estimate	Std. Error	z value	P-value
Model: All avian				
Intercept	1.24163	0.18534	6.699	<0.001
Habitat Diversity	-0.6684	0.29011	-2.304	<0.05
Livestock-keeping	-0.23911	0.12504	-1.912	0.056
Avian Species Richness:Fruit/Nectar	0.03821	0.01097	3.482	<0.001
Avian Species Richness:Invertebrate	0.02824	0.01117	2,528	<0.05
Avian Species Richness:Omnivore	0.03247	0.01094	2.967	<0.01
Avian Species Richness:Seedeater	0.01586	0.01067	1.487	0.137
Model: Seed-eating birds only				
Intercept	1.2886	0.1603	8.039	<0.001
Habitat Diversity	-0.7249	0.3142	-2.307	<0.05
Model: Seed-eating birds only (spatial)				
Intercept	0.91819	0.05918	15.514	<0.001
MEM2	0.13928	0.05531	2.518	<0.05
MEM6	0.11279	0.05513	2.046	<0.05
MEM10	0.12502	0.04948	2.526	<0.05
MEM13	-0.13317	0.05271	-2.526	<0.05
MEM16	-0.09168	0.05057	-1.813	0.07

Table 6.1. Estimated regression parameters, standard errors, *z*-values and p-values for the Poisson generalised linear mixed effects models (GLMMs) presented in eqns (1), (2) and (3). In these models, virulence gene diversity in avian-borne *E. coli* is related to different ecological and household determinants. Moran's eigenvector maps (MEMs) indicate the spatial scales across which variation in MDR carriage or profile length occurs. Increasing MEM numbering indicates increasingly finer spatial scales.

The effect of urban change on the diversity of genes encoding antimicrobial resistance (AMR) was investigated in a similar way, utilising the same set of isolates and explanatory variables as used for virulence genes. However, due to zero-inflation and overdispersion, a zero-inflated negative-binomial "type 1" generalised linear model (GLM) was fitted in the R package 'glmmADMB' [108]. Intercept-only models indicated that household and sublocation did not explain any variance in the dataset, and as such were not included in the full model. Backwards step-wise selection resulted in *livestock-keeping* (categorical) and *human density* (continuous) being retained as fixed effects with an interaction term, in the following model:

$$\text{AMR gene diversity}_{i\ i} = \text{livestock.keep:human density}_i$$

(eqn 4)

The model was assessed by checking for overdispersion (dispersion statistic: 0.68), heterogeneity, residual, and non-linear patterns, none of which did not indicate any problems. α-diversity of AMR genes was significantly associated with increasing human density, but only in livestock-keeping households (Table 6.2). To test whether

this interaction was dependent upon host functional-group membership, the optimal model in eqn 4 was fitted independently for isolates derived from seed-eating and non-seed-eating birds. This indicated that the relationship between AMR gene diversity, livestock keeping and human density was only present for seed-eating birds (Figure 6.2). To explore this relationship further, the fixed covariate *livestock-keeping* in eqn 4 was replaced with *livestock density* (correlation prevented both from being fitted in the same model). The resulting model and interaction term were also significant, suggesting that increases in human density at higher densities of livestock result in increased diversity of AMR genes in seed-eating birds (Table 6.2).

Finally, a spatial zero-inflated negative binomial "type 1" model was built for α-diversity of AMR genes in isolates derived from seed-eating birds. Although three dbMEMs were retained in the final model, from a total of 17 dbMEMs modelling positive spatial structure amongst the isolates, none of the fixed covariates were significantly associated with AMR gene diversity.

$$\text{AMR gene diversity}_i = \text{MEM1}_i + \text{MEM13}_i + \text{MEM17}_i \quad \text{(eqn 5)}$$

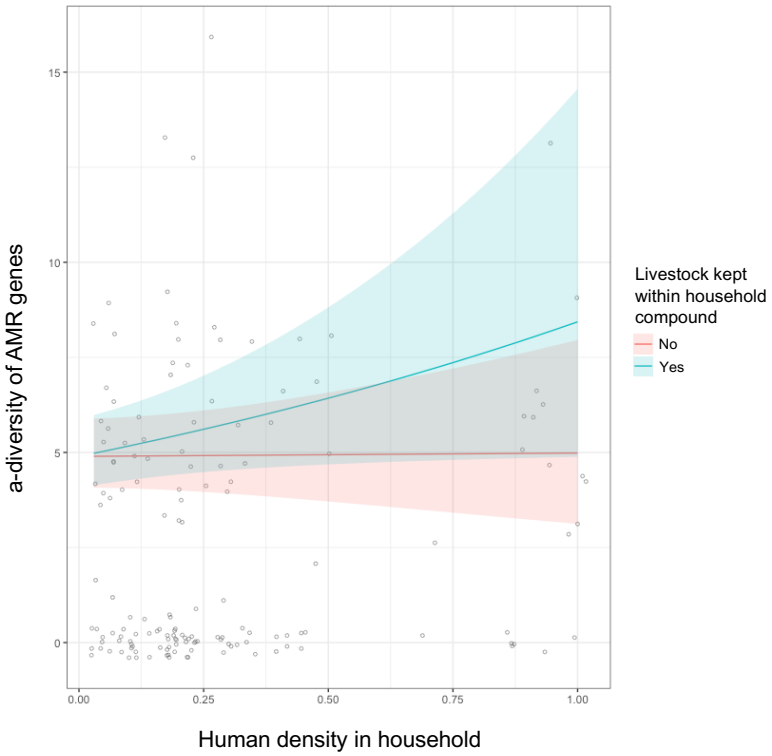


Figure 6.2. Fit of the zero-inflated negative binomial 'type 1' generalised linear model (GLM) shown in eqn (4), applied to isolates from seed-eating birds. This model demonstrates that when livestock were part of the vertebrate host community at household interfaces, α-diversity of AMR genes in birds increased with human density. Shading on either side of each line represents 95% confidence intervals. Grey points represent data points for individual seed-eating birds.

Model terms	Estimate	Std. Error	z value	P-value
Model: All avian				
Intercept	1.6253	0.0716	22.7	<0.001
Non Livestock-keeping:Human Density	-0.0238	0.1915	-0.12	0.901
Livestock-keeping:Human Density	0.4981	0.2286	2.18	<0.05
Model: Seed-eating birds only (1)				
Intercept	1.5885	0.0937	16.95	<0.001
Non Livestock-keeping:Human Density	0.0177	0.2153	0.08	0.934
Livestock-keeping:Human Density	0.5438	0.2625	2.07	<0.05
Model: Seed-eating birds only (2)				
Intercept	1.6321	0.0704	23.18	<0.001
Livestock Density:Human Density	0.8635	0.4071	2.12	<0.05
Model: Seedeating birds only (spatial)				
Intercept	1.6627	0.0625	26.62	<0.001
MEM1	-0.1061	0.0645	-1.65	0.1
MEM13	-0.1272	0.0666	-1.91	0.056
MEM17	-0.1484	0.0773	-1.92	0.055

Table 6.2. Estimated regression parameters, standard errors, z-values and P-values for the Poisson generalised linear mixed effects models (GLMMs) presented in eqns (4) and (5). In these models, the diversity of antimicrobial resistance genes present in avian-borne *E. coli* is related to different household determinants. Moran's eigenvector maps (MEMs) indicate the spatial scales across which variation in MDR carriage or profile length occurs. Increasing MEM numbering indicates increasingly fine spatial scales.

6.4.2 Genetic structure of *E. coli* populations in wildlife, and in relation to livestock and humans across Nairobi – isolation by environment and isolation by distance

cgMLST schemes were determined for a total of 1092 gene loci, in 1485 *E. coli* whole genome sequences (wildlife, n=413; livestock, n=568; humans, n=293; outside environment, n=211). DAPC analysis performed on avian, rodent and bat isolates without prior groupings generated an inconclusive estimate of $K = 6$ genetic clusters along the first 111 retained principal components (PCs) (suppl. Figure 6.2a). Cross-validation procedures and α-score optimisation suggested that roughly 111 PCs should be retained in the analysis (suppl. Figure 6.2b). Plotted geographically, all six genetic clusters were spatially admixed (Figure 6.3). When DAPC was performed using *a-priori* selected wildlife functional groups, genetic differentiation was evident between functional groups along the first two discriminant axes; fruit/nectar-eating birds and bats were discriminated from other functional groups by axis 1, whilst rodents were discriminated from other functional groups by axis 2 (Figure 6.4a). Substantial admixture was evident between all other wildlife functional groups. When *a-priori* wildlife, livestock, human and environmental groups were specified, cross-

validation and α-score procedures suggested that in the region of 200 PCs should be retained for the analysis (suppl. Figure 6.2c). As previously, axis 1 discriminated fruit/nectar-eating birds and bats from all other source groups, whilst axis 2 showed considerable admixture between all other source groups (Figure 6.4b). Of the source groups on axis 2, those most distinct from each other were humans and scavenging birds, and rodents, goats and insectivorous bats.

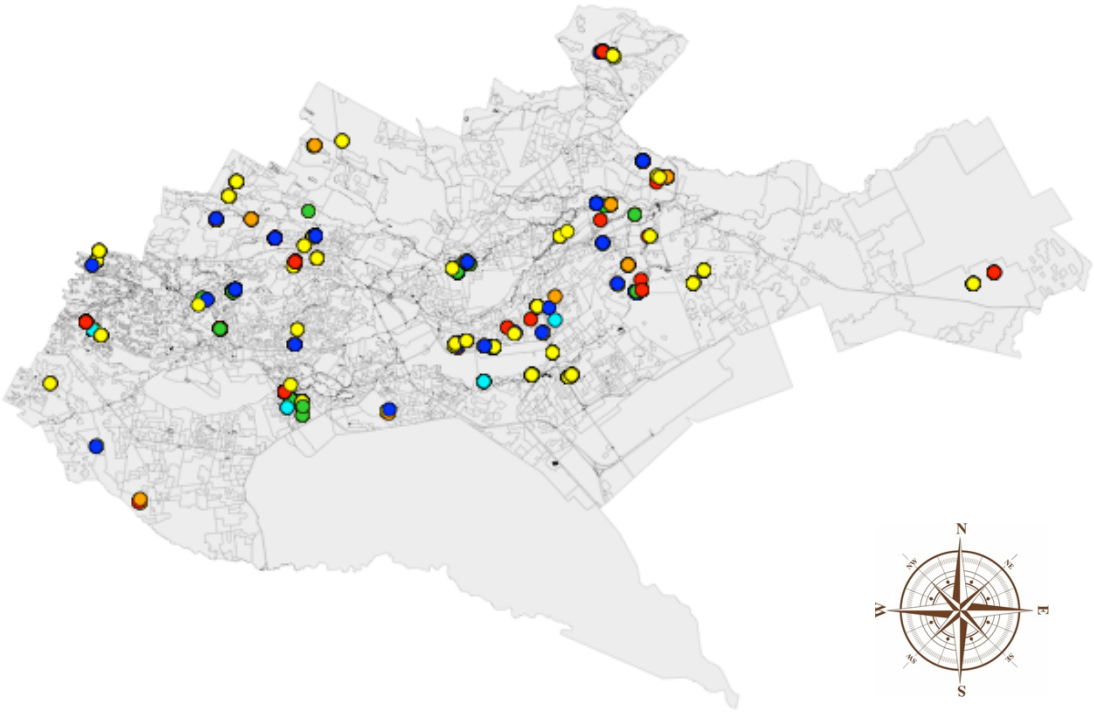


Figure 6.3. Map demonstrating spatial admixture of *K* = 6 genetic groups of *E. coli*, as inferred by DAPC and k-means clustering performed on all wildlife-borne *E. coli* isolates. Colours (yellow, red, blue, green, orange and turquoise) represent genetic groups.

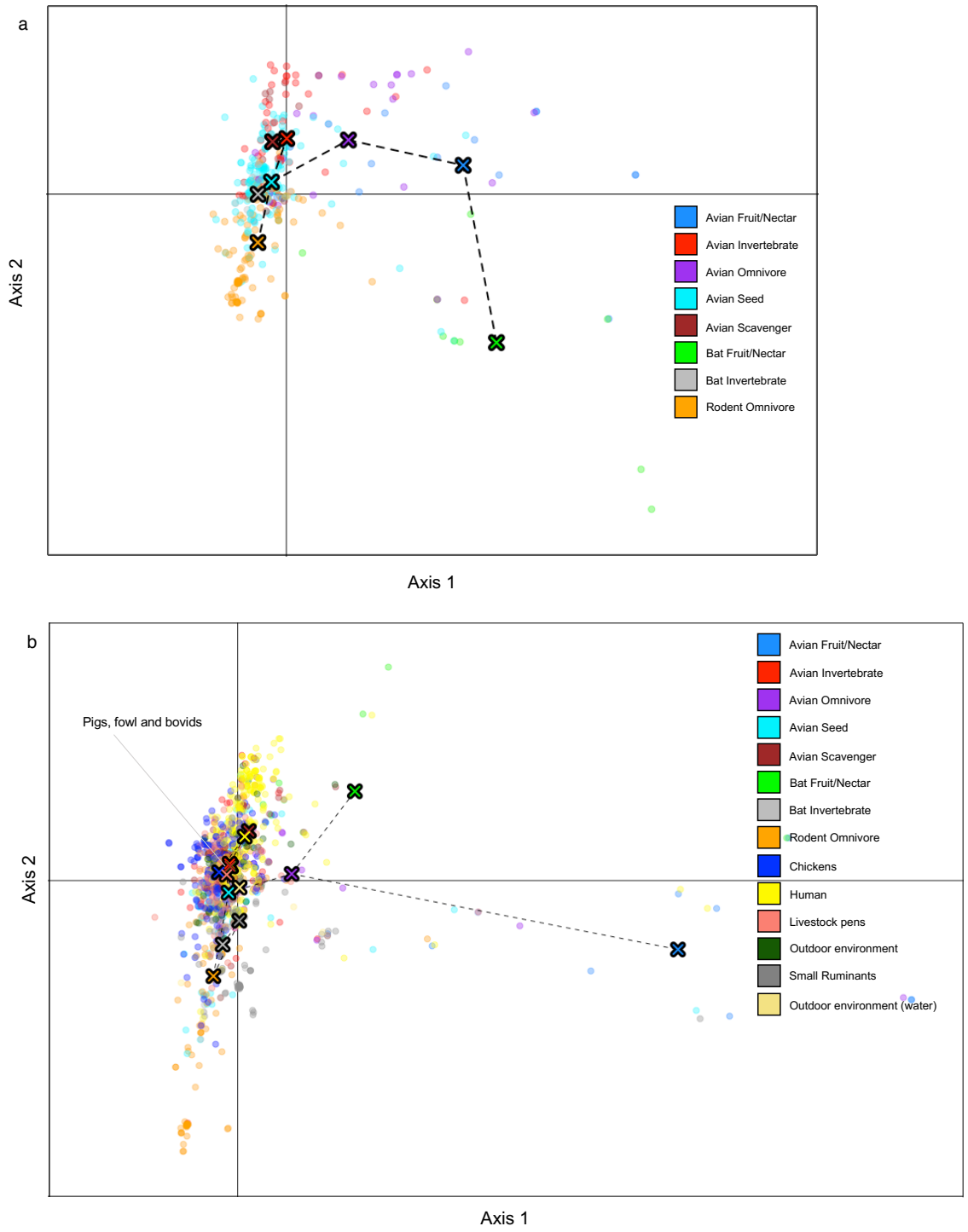


Figure 6.4. Scatterplots of discriminant analysis of principal components (DAPC), showing the first two discriminant axes for *E. coli* population genetic differentiation of **a)** wildlife functional groups chosen *a-priori*, and **b)** wildlife functional groups, livestock taxa, humans and environmental samples chosen *a-priori*. Colours represent taxa chosen *a-priori*. Dashed black line represents minimum-spanning-tree, based on the squared distances between populations. Crosses indicate the center of each functional group. Centroids for pigs, fowl and bovids are obscured on the plot, and are thus indicated directly by the grey line.

Differentiation between hosts within taxonomic functional groups accounted for 97% of genetic variance explained by AMOVA, and significant genetic differentiation was not present between functional groups within households, or between wildlife in different households (Table 6.3). This suggests that *E. coli* is shared across wildlife taxa within households, and between wildlife in different households. Comparing pairwise F_{ST} between wildlife functional groups and other epidemiological compartments, the following patterns emerged (suppl. Table 6.2, Figure 6.5). Fruit/nectar-eating birds and bats had consistently higher pairwise F_{ST} values, indicating that they were genetically differentiated from other source groups, and supporting the results of the DAPC analysis. All other wildlife demonstrated considerable admixture with populations of livestock and humans, and the environment ($F_{ST} < 0.05$). Seed - eating birds were minimally differentiated from humans, cattle and the household environment ($F_{ST} <0.01$), insectivorous bats were minimally differentiated from cattle and small ruminants ($F_{ST} < 0.004$), and scavenging birds showed no measurable differentiation from humans, poultry, pigs or environmental samples ($F_{ST} = 0$).

Stratification	Df	Sum Sq	Mean Sq	Variance components	% total variance	ϕ - statistics	p-value
Between households	87	41243.400	474.062	1.471	0.328	0.028	0.341
Between functional groups, within households	59	27051.670	458.503	11.116	2.481	0.025	0.124
Between hosts, within functional groups	231	100608.410	435.534	435.534	97.191	0.003	0.001

Table 6.3. Results of analysis of molecular variance (AMOVA), depicting partitioning of genetic variation of *E. coli* between wildlife hosts and their population within- and between- households.

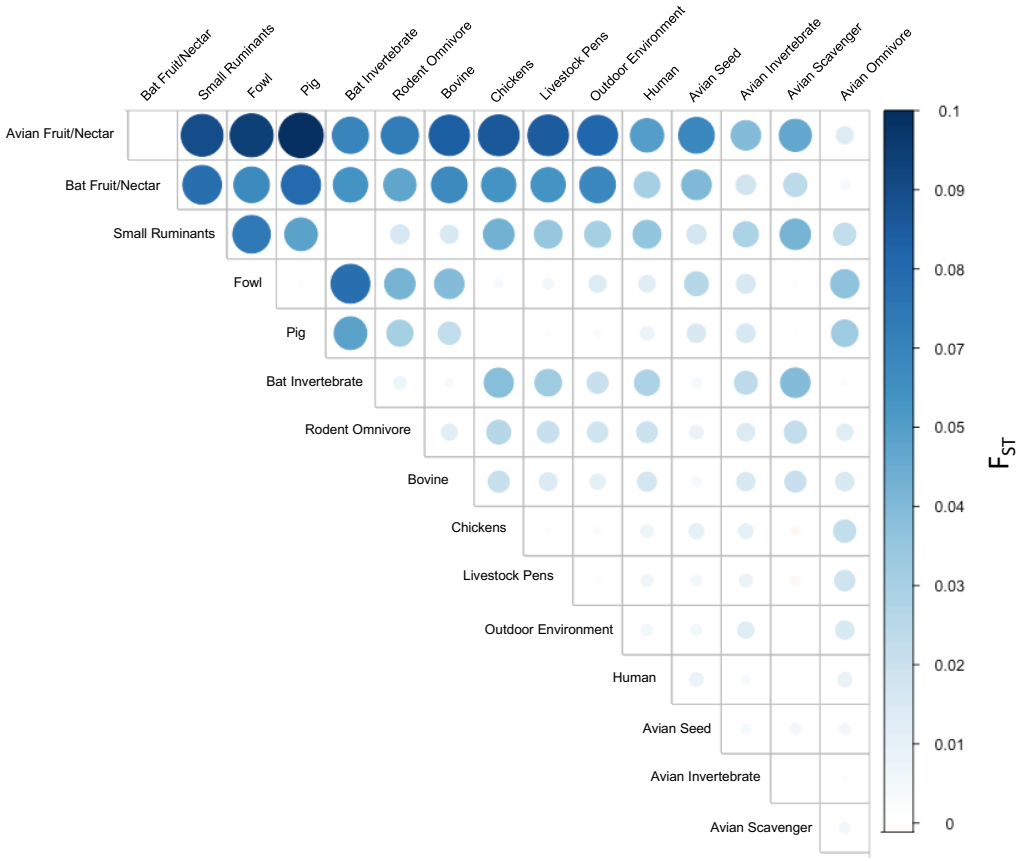


Figure 6.5. Heat map/bubble map showing pairwise Weir and Cockerham F_{ST} values, between wildlife functional groups, livestock taxa, humans and environmental groups. F_{ST} is measured on a scale of 0 to 1 (minimally and maximally divergent, respectively). Colour intensity and size of points indicates larger FST values.

The relationship of wildlife to other epidemiological compartments was further explored using distance-based redundancy analysis (db-RDA). This indicated that both host/source identity and the spatial scale at which samples were collected had a statistically significant influence on genetic dissimilarity, and thus differentiation of *E. coli*. When sequence data from all sources was considered, the optimal source db-RDA model explained considerably more variance in genetic (dis)similarity between isolates than the spatial model (R^2_{adj} = 0.014 and 0.001 respectively), and showed that genetic (dis)similarity was only determined at a very fine scale (MEM33, P<0.05). Genetic (dis)similarity also varied significantly between sources (P=0.001) (Table 6.4). On visual inspection of the source model db-RDA triplot, fruit/nectar–eating birds and bats explained most variance, and were correlated, suggesting that they shared similar cgMLST profiles (Figure 6.6). Correlative relationships between other sources in the db-RDA triplot broadly overlap with those in the equivalent DAPC plot, and the pairwise F_{ST} plot (Figures 6.5 & 6.6).

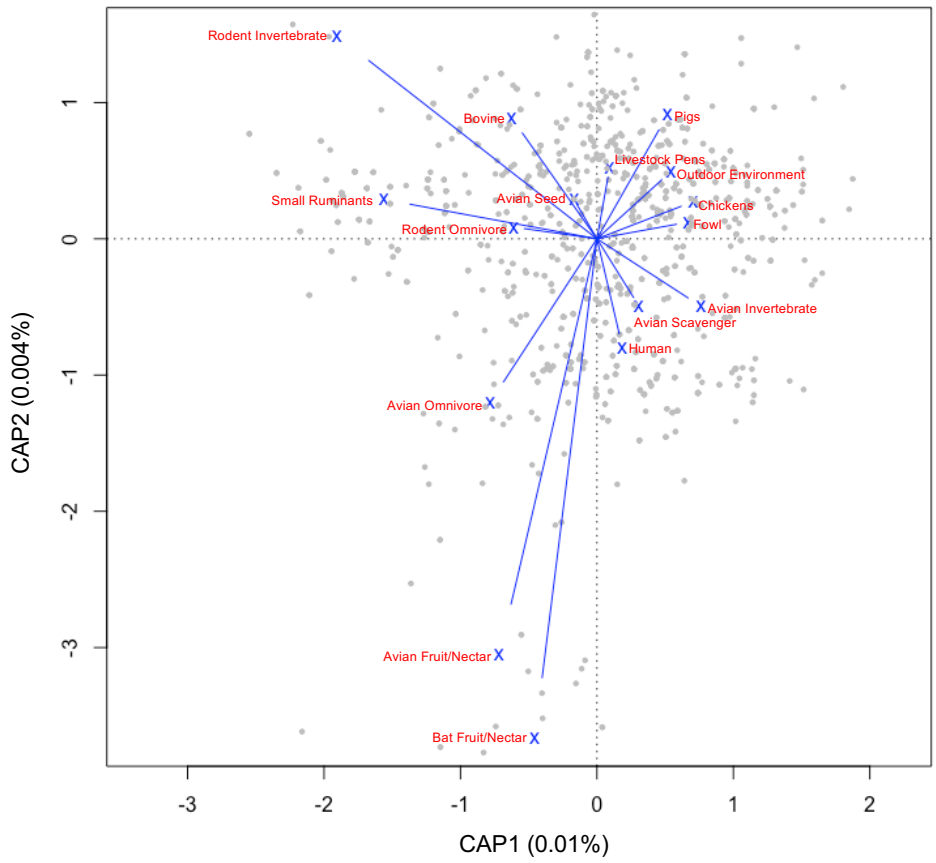


Figure 6.6. Correlation triplot of optimal distance-based redundancy analysis (db-RDA) model for the effect of host/source on genetic (dis)similarity between *E. coli* isolates. Response variables (principle coordinates representing the Jaccard (dis)similarity between *E. coli* isolates) are represented as grey points, and levels of the explanatory variable are blue lines, crosses (centroids) and red text. The first and second canonical axes explained 0.01% and 0.004% of total variance in the response distance matrix.

	Spatial Models				**Host/Source Models**		
	Wildlife	**Livestock**	**Environment**	**All**	**Wildlife**	**Livestock**	**All**
R^2_{adj}	*0.004*	*0.006*	*0.016*	*0.001*	*0.01*	*0.016*	*0.014*
Spatial Variables	Total = 30	Total = 16	Total = 20	Total = 38	Total = 30	Total = 16	Total = 38
MEM3	-	-	**0.023**	-	PARTITIONED OUT	PARTITIONED OUT	PARTITIONED OUT
MEM4	-	**0.021**	-	-			
MEM7	-	-	**0.032**	-			
MEM8	-	-	**0.04**	-			
MEM9	**0.044**	-	**0.028**	-			
MEM10	-	**0.014**	-	-			
MEM11	**0.025**	-	-	-			
MEM16	-	**0.009**	-	-			
MEM21	**0.025**	-	-	-			
MEM33	-	-	-	**0.022**			
Host/Source							
	PARTITIONED OUT				**0.001**	**0.001**	**0.001**

Table 6.4. R^2_{adj} values for optimal distance-based redundancy analysis (db-RDA) models examining the effects of spatial scales and host/source on genetic (dis)similarity between *E. coli* isolates, and p-values (in bold) for variables included in these models. Dashes represent variables not present in each optimal model.

To investigate whether the influence of spatial scale on genetic structure differed between wildlife isolates and those recovered from other source groups (livestock, humans and the environment), separate db-RDA models were built for isolates belonging to each source group (Table 6.4). The optimal model for wildlife isolates showed that significant genetic differentiation occurred at intermediate (MEM9 and MEM11) and fine (MEM21, from a total of 30 MEMs carried forward) spatial scales across the city ($P<0.01$, $R^2_{adj} = 0.004$). For livestock, the optimal model only included fine, intermediate and broad spatial scales (MEM4, MEM10 and MEM16, out of a total of 16 MEMs carried forward) ($P=0.001$, $R^2_{adj} = 0.006$), whilst for environmental samples, genetic differentiation occurred at broad (MEM3) and intermediate (MEM7, MEM8 and MEM9 out of a total of 20 MEMs carried forward) scales ($P<0.01$, $R^2_{adj} = 0.016$). The full model including all 22 MEMs modelling positive spatial structure amongst humans in the study was not statistically significant, and as such there was no evidence of spatial structure in the genetic differentiation of *E. coli* in humans. Separate models run for wildlife and livestock isolates showed that functional/taxonomic groups determined small but statistically significant proportions of the variance in genetic (dis)similarity (Wildife: $P=0.001$, $R^2_{adj} = 0.01$; Livestock: $P<0.001$, $R^2_{adj} = 0.016$) (Table 6.4). That each canonical model in these analyses accounted for less than 2% of variance in the genetic distance between isolates, suggests that un-measured determinants (likely at the individual host level) are driving most of the genetic differentiation in *E. coli*.

6.5 Discussion

In this chapter, high resolution genetic data was used to explore in greater detail key inferences and hypotheses generated from the first two data chapters, assessing the response of bacterial genetic communities to urban land-use change, and investigating the population structure of *E. coli* between hosts. Deterministic forces (both ecological and anthropogenic) operating across the urban landscape of Nairobi influence the structure of bacterial genetic communities within vertebrate host communities. The results of this study are discussed below, beginning with the effects of urban land-use change on the diversity of virulence and AMR genes.

6.5.1 Response of mobile genetic communities to land-use change

The influence of biodiversity on the diversity of parasites hosted in vertebrate communities is a fundamental question relating to the role of wildlife in the emergence of zoonotic diseases in humans. In a novel analysis using genetic markers that are exchanged horizontally between *E. coli* as a proxy for the relationship between parasites and their hosts, this question was explored in relation to urban land-use change. MGE's are exchanged between microorganisms in a manner that resembles transmission of directly transmitted ubiquitous parasites between their hosts.

For virulence genes, the richness of host communities was responsible for shaping the diversity of genes present in *E. coli* isolates, with increases in avian diversity leading to a higher diversity of virulence genes within their *E. coli*. This follows an expected pattern for communities of hosts and their parasites. Assuming each vertebrate host harbours at least some *E. coli* bearing unique virulence genes, increasing vertebrate species diversity will increase the diversity of virulence genes circulating in the population [46] (reviewed in [157]). As shown in this study system, increased vertebrate diversity results in *E. coli* (and by extension, avian hosts) acquiring a greater diversity of *E. coli*–borne virulence genes, because of exposure to a larger 'pool' of available genes in the vertebrate host community. The composition and size of this 'pool' would be hypothesised to vary across a gradient of urban land-use change, as the structure of avian communities change in response to the changes in habitat structure and biotic resource provision described in Chapter 4. The results of this study also indicate that differences in the response of wildlife species to urban land-use change could play a part in determining whether microbial genetic diversity maps to host community diversity or not. For example, the diversity of virulence genes in *E. coli* derived from seed-eating birds, which show more synanthropic behaviour than other functional groups (see Chapter 4), was not predicted by avian community diversity, instead being influenced by changes in biotic habitat diversity - increasing virulence gene diversity was linked to decreasing biotic complexity of habitats. Further evidence for the role of host taxa in shaping the response of microbial genetics to urban land-use change was evident when considering genetic determinants of antimicrobial resistance in *E. coli*. Increasing human and livestock density were the main determinants for higher AMR gene diversity in avian-borne *E. coli*, but this only

applied to isolates recovered from seed-eating birds. As demonstrated in Chapter 5, carriage of phenotypic AMR-*E. coli* present in seed-eating birds was also linked to livestock and humans within households, and taken together these results underline the importance of this functional group of birds in AMR epidemiology. Importantly, the relationship between AMR gene diversity and human density only existed in the presence of livestock, providing further evidence that households act as an interface for the transmission of genes encoding AMR from livestock to wild birds. Livestock and human density, both identified in Chapter 4 as highly correlated factors that differentiate the structure of wildlife-livestock-interfaces across the urban landscape, were therefore responsible for influencing the diversity (or 'pool') of AMR genes present and/or promoting contact with synanthropic wildlife. This likely resulted in spillover of bacteria and/or their genetic elements from livestock to wildlife at household interfaces.

These results are important for several reasons. Firstly, they demonstrate the presence of two opposing epidemiological gradients across the urban landscape, in which mobile communities of microbial genes respond to changes in the richness and density of vertebrate host communities (which may be confounded by the ecological traits of the host within which that organism resides) (Figure 6.7a). Although the horizontal exchange mechanisms involved in the transfer of these genes are unlikely to directly mimic the dynamics of transmission within parasite communities, such deterministic patterns could also be displayed by communities of parasites subject to the same changes in host community structure. For example, abundance of hosts has been linked to parasite species richness in a number of previous studies [231,232], and increasing diversity of helminth parasitism in Southeast Asian murids has been positively correlated with a gradient of anthropogenic habitat change [54]. Secondly, these results provide clear evidence for a mechanism by which anthropogenic processes tied to urban land-use change result in spillover of microbes and/or their genes between vertebrate host compartments at wildlife-livestock-human interfaces (Figure 6.7b). A pathway of epidemiological connectivity exists between synanthropic birds and livestock, across which parasites could be transmitted. Thirdly, these results show that ecological and anthropogenic processes associated with urban land-use change may simultaneously exert very different genetic selection pressures on the same species of parasite, with important implications for public health. For bacterial organisms such as *E. coli*, exposure to larger pools of genetic diversity that promote

uptake and fixing of virulence and AMR genes can confer adaptive advantages such as virulence and drug resistance, whilst exposure of viruses to genetic diversity can play an important role in their evolution (e.g. influenza A viruses [233]). Divergence associated with horizontal gene transfer between closely related microbial strains can lead to the emergence of novel pathogens [234,235].

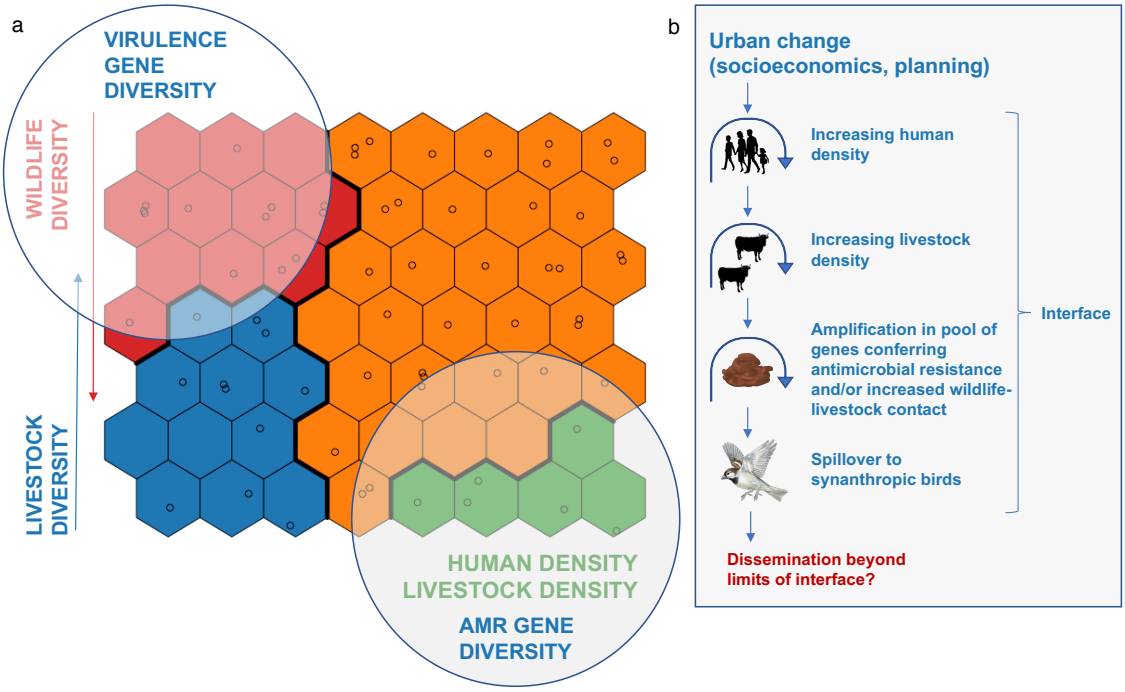


Figure 6.7. **a)** Diagrammatic representation of how epidemiological gradients in α-diversity of avian-borne *E. coli* virulence and antimicrobial resistance (AMR) genes overlay on broad-scale trends in host community characteristics at urban household interfaces. The characteristics of household host communities are represented in the form of a Self-organising map, adapted from Chapter 1, in which groups of coloured hexagons (red, blue, orange and green) indicate the membership of households to clusters with similar host community structures. Avian virulence gene diversity is highest in households with more avian taxonomic diversity, whilst AMR gene diversity is highest in households with higher livestock and human density. **b)** Schematic illustrating the sequence of events leading to spillover of AMR genetic determinants between livestock and synanthropic birds at household interfaces.

6.5.2 Population structure of *E. coli* in wildlife

Most of the core-genome genetic variation originating in populations of wildlife-borne *E. coli* was found between individual hosts, with very little differentiation existing between taxonomic functional groups within households, or between populations of wildlife in different households. This population structure is typical of *E. coli*, and it is common for most of the genetic differentiation in these

bacteria to be explained by variation within or between individual hosts [205,236]. As such, transmission of *E. coli* between wildlife is not restricted to functional groups at household interfaces. Once the confounding effects of host taxa were removed, spatial differentiation between *E. coli* in wildlife was small, and only represented at fine and intermediate spatial scales. With these results, a hypothesis of population differentiation through 'isolation by distance' could be rejected; the presence of minimal intermediate spatial structure, and absence of broad-scale structure, suggests that gene flow occurs either directly or indirectly between remote populations of wildlife across the urban landscape. In comparison, *E. coli* derived from environment samples showed considerably more spatial differentiation, at intermediate and broad scales, suggesting that gene flow between environmental sources was more restricted. Like wildlife, *E. coli* from livestock showed structure across fine and intermediate scales, whilst human-borne *E.coli* did not show any spatial differentiation, suggesting gene flow between humans at all scales across the city.

At a city-wide scale, the main determinant of *E. coli* genetic differentiation was the feeding ecology of wildlife hosts. Wildlife occupying extreme dietary niches (diets based on fruit/nectar) had clearly differentiated populations of *E. coli* from those with different feeding strategies, which were all relatively admixed. Feeding ecology was also responsible for differentiating wildlife-borne *E. coli* from isolates originating from livestock, humans and the environment. Apart from fruit/nectar-eating birds and bats, *E. coli* from wildlife showed considerable admixture with populations from livestock and humans, and the environment, making it difficult to assess genetic differentiation between them. However, seed-eating and scavenging birds showed the least genetic differentiation from humans, certain livestock species, and the environment. This, given that the feeding ecology of these functional groups bring them into close contact with the ground (and in the case of scavenging birds, human waste), makes intuitive sense. Although the effects of diet and enteric selection on *E. coli* could not be separated from exposure through foraging, these results support inferences made earlier in this chapter and in Chapter 5, which linked foraging-related life-history traits with exposure to phenotypic and genotypic AMR, suggesting that synanthropic foraging ecology in wildlife plays an important role in host exposure to microorganisms.

In this chapter, high resolution genetic data collected as part of a structured epidemiological study, was utilised to study bacterial epidemiology in a multi-host

urban system. The scale of sampling conducted in this study (representing sympatric wildlife, livestock and human communities along a gradient of urban land-use) has not been previously attempted. Whilst this provided the opportunity to explore novel hypotheses that, until recently could not have been tested, this dataset is not without epidemiological limitations, and the results presented in this chapter should be interpreted with the following considerations in mind. To fully characterise MGE diversity, and the form of MGE transfer occurring between hosts, long-read sequencing (e.g. using PacBio) would have been required. Lower resolution short read WGS, as used in this study, results in less accurate identification of MGEs (particularly those borne on plasmids) with implications for epidemiological inference [237]. Whether this would have had a significant effect on these results is unclear, since the focus of this study was on patterns of diversity rather than characterising individual MGEs. The hierarchical design of the study, including two levels of nestedness, was accounted for in our models, but the number of replicated household sampling units (99) was relatively small. Along with substantial variation in sample size between host taxa, this has undoubtedly introduced a level of uncertainty to the statistical tests and modelling conducted on this data, making it less likely that a true effect is detected when present. The sensitivity of commensal *E. coli* in identifying transmission pathways for other pathogens should also be considered with caution. Differences in characteristics (such as shedding rates and effects on host behaviour) between commensal and pathogenic organisms may have epidemiological consequences that reduce their representation of one another. In addition, by only sequencing a single *E. coli* isolate from each host, the within-host genetic diversity of *E. coli* was not considered. Previous molecular studies on *E. coli* (albeit it in different hosts, and using lower resolution sequencing technology), have demonstrated considerable within-host diversity across vertebrate taxa [80,205,238]. However, the decision to sequence a single isolate from each host was made as a necessary, cost-based trade-off between genetic resolution and sample size. The effects of restricted sample size would only act to increase Type II error in our results (i.e. conservative statistical inference, or missed signal in the data), and are thus unlikely to affect the validity of our findings.

6.6 Conclusion

Molecular epidemiology of wildlife-borne *E. coli* at household interfaces is linked to structural changes in sympatric wildlife, livestock and human populations. In demonstrating that it is possible to link epidemiological processes in wildlife to their drivers across urban landscapes at multiple scales, this study has taken the first step towards forecasting the effects of urban land-use change on disease emergence within a developing city. This justifies the detailed approach to studying wildlife-livestock-human interfaces advocated throughout this thesis, where efforts to fully characterise the ecological and anthropogenic responses to urban change (undertaken in Chapter 4) provided a set of variables that could be used to test epidemiological hypotheses across different scales. More comprehensive scrutiny of this dataset, extending analysis of MGE diversity to humans and livestock, would provide valuable insight into the epidemiological responses of these compartments to land-use change, but was outside the scope of this thesis. By considering MGE diversity in a single species of *Enterobacteriaceae* as a proxy for parasite diversity, this study has necessarily taken a reductionist approach to address important hypotheses that otherwise could not have been answered using this dataset. The limitations in using a model organism such as *E. coli* could be addressed through metagenomic methods, that could be used to characterise the structural response of parasite communities to the environmental drivers of land-use change. Such methods will be required to understand how changes in parasite diversity, and the uptake and fixing of genes by parasites, translate to emergence and manifestation of clinical disease in wildlife, livestock and humans.

6.7 Supplementary Data

Gene	Gene	Gene	Gene	Gene	Gene	Gene	Gene	Gene	Gene	Gene	Gene	Gene	Gene	Gene	Gene	Gene	Gene	Gene	Gene	Gene	Gene
C_RS00200	C_RS00015	C_RS00020	C_RS00025	C_RS00040	C_RS00045	C_RS00050	C_RS00055	C_RS00065	C_RS00070	C_RS00120	C_RS00125	C_RS00130	C_RS00135	C_RS00140	C_RS00145	C_RS00150	C_RS00160	C_RS00170	C_RS00185	C_RS00190
C_RS00470	C_RS00210	C_RS00215	C_RS00220	C_RS00225	C_RS00245	C_RS00250	C_RS00255	C_RS00265	C_RS00270	C_RS00285	C_RS00290	C_RS00295	C_RS00300	C_RS00305	C_RS00315	C_RS00320	C_RS26070	C_RS00455	C_RS00460	C_RS00465
C_RS00665	C_RS00480	C_RS00490	C_RS00495	C_RS00500	C_RS00505	C_RS00515	C_RS00520	C_RS00525	C_RS00530	C_RS00535	C_RS00540	C_RS00545	C_RS00565	C_RS00570	C_RS00575	C_RS00580	C_RS00605	C_RS00610	C_RS00615	C_RS00620
C_RS00880	C_RS00670	C_RS00680	C_RS00695	C_RS00700	C_RS00705	C_RS00710	C_RS00715	C_RS00725	C_RS00730	C_RS00735	C_RS00740	C_RS00745	C_RS00750	C_RS00755	C_RS00770	C_RS00820	C_RS00825	C_RS00835	C_RS00845	C_RS00875
C_RS01005	C_RS00885	C_RS00890	C_RS00895	C_RS00905	C_RS00910	C_RS00925	C_RS00930	C_RS00935	C_RS00940	C_RS00945	C_RS00950	C_RS00955	C_RS00960	C_RS00970	C_RS00975	C_RS00980	C_RS00985	C_RS00990	C_RS00995	C_RS01000
C_RS02280	C_RS01010	C_RS01015	C_RS01020	C_RS01025	C_RS01030	C_RS01035	C_RS01045	C_RS01055	C_RS01065	C_RS01160	C_RS01170	C_RS01175	C_RS01180	C_RS01185	C_RS01190	C_RS02160	C_RS02165	C_RS02210	C_RS02270	C_RS02275
C_RS02430	C_RS02290	C_RS02295	C_RS02300	C_RS02305	C_RS02310	C_RS02320	C_RS02325	C_RS02335	C_RS02340	C_RS02345	C_RS02350	C_RS02355	C_RS02360	C_RS02370	C_RS02375	C_RS02390	C_RS02395	C_RS02400	C_RS02415	C_RS02425
C_RS02575	C_RS02445	C_RS02450	C_RS02455	C_RS02460	C_RS02465	C_RS02475	C_RS02480	C_RS02490	C_RS02495	C_RS02505	C_RS02515	C_RS02520	C_RS02525	C_RS02535	C_RS02540	C_RS02545	C_RS02550	C_RS02555	C_RS02560	C_RS02565
C_RS02945	C_RS02625	C_RS02630	C_RS02645	C_RS02655	C_RS02660	C_RS02670	C_RS02680	C_RS02695	C_RS02700	C_RS02705	C_RS02710	C_RS02715	C_RS02720	C_RS02730	C_RS02735	C_RS02810	C_RS02815	C_RS02925	C_RS02935	C_RS02940
C_RS09560	C_RS08630	C_RS08635	C_RS09375	C_RS09380	C_RS09400	C_RS09405	C_RS09410	C_RS09415	C_RS09440	C_RS09445	C_RS09450	C_RS09455	C_RS09460	C_RS09510	C_RS26950	C_RS09525	C_RS09530	C_RS09545	C_RS09550	C_RS09555
C_RS09685	C_RS09565	C_RS09570	C_RS09575	C_RS09580	C_RS09585	C_RS09590	C_RS09595	C_RS09600	C_RS09605	C_RS09610	C_RS09615	C_RS09630	C_RS09640	C_RS09645	C_RS09650	C_RS09655	C_RS09660	C_RS09670	C_RS09675	C_RS09680
C_RS09995	C_RS09715	C_RS09785	C_RS09790	C_RS09795	C_RS09800	C_RS09805	C_RS09810	C_RS09835	C_RS09840	C_RS09870	C_RS09920	C_RS09930	C_RS09935	C_RS09940	C_RS09945	C_RS09960	C_RS09965	C_RS09970	C_RS09975	C_RS09985
C_RS10330	C_RS10000	C_RS10005	C_RS10010	C_RS10100	C_RS10105	C_RS10110	C_RS10145	C_RS10155	C_RS10160	C_RS10180	C_RS10205	C_RS10210	C_RS10220	C_RS10245	C_RS10255	C_RS10265	C_RS10270	C_RS10275	C_RS10320	C_RS10325
C_RS10535	C_RS10335	C_RS10340	C_RS10375	C_RS10395	C_RS10400	C_RS10410	C_RS28605	C_RS10445	C_RS10450	C_RS10455	C_RS10460	C_RS10470	C_RS10485	C_RS10490	C_RS10495	C_RS10500	C_RS10505	C_RS10510	C_RS10525	C_RS10530
C_RS10715	C_RS10545	C_RS10550	C_RS10560	C_RS10570	C_RS10575	C_RS10580	C_RS10595	C_RS10600	C_RS10610	C_RS10620	C_RS10625	C_RS10635	C_RS10640	C_RS10650	C_RS10655	C_RS10660	C_RS10670	C_RS10675	C_RS10690	C_RS10705
C_RS10850	C_RS10720	C_RS10725	C_RS10735	C_RS10740	C_RS10745	C_RS10750	C_RS10755	C_RS10760	C_RS10765	C_RS10770	C_RS10775	C_RS10785	C_RS10790	C_RS10800	C_RS10805	C_RS10815	C_RS10820	C_RS10825	C_RS10840	C_RS10845
C_RS12325	C_RS10930	C_RS10960	C_RS10965	C_RS28625	C_RS10975	C_RS10980	C_RS10985	C_RS10990	C_RS10995	C_RS11015	C_RS11465	C_RS12055	C_RS12060	C_RS12065	C_RS12085	C_RS12285	C_RS12290	C_RS12295	C_RS12300	C_RS12310
C_RS12645	C_RS12330	C_RS12335	C_RS12340	C_RS12345	C_RS12355	C_RS12360	C_RS12445	C_RS12450	C_RS12460	C_RS12470	C_RS12475	C_RS12480	C_RS12530	C_RS12585	C_RS12590	C_RS12595	C_RS12610	C_RS12630	C_RS12635	C_RS12640
C_RS12900	C_RS12650	C_RS12675	C_RS12680	C_RS12685	C_RS12695	C_RS12720	C_RS12725	C_RS12735	C_RS12740	C_RS12750	C_RS12770	C_RS12775	C_RS12780	C_RS12815	C_RS12860	C_RS12865	C_RS12875	C_RS12880	C_RS12890	C_RS12895
C_RS13275	C_RS12905	C_RS12910	C_RS12915	C_RS12920	C_RS12925	C_RS12960	C_RS12965	C_RS12970	C_RS13160	C_RS13175	C_RS13180	C_RS13185	C_RS13190	C_RS13195	C_RS13205	C_RS13215	C_RS13250	C_RS13255	C_RS13265	C_RS13270
C_RS13495	C_RS13280	C_RS13285	C_RS13290	C_RS13295	C_RS13305	C_RS13310	C_RS13315	C_RS13320	C_RS13330	C_RS13335	C_RS13435	C_RS13440	C_RS13445	C_RS13450	C_RS13455	C_RS13460	C_RS13475	C_RS13480	C_RS13485	C_RS13490
C_RS13630	C_RS13500	C_RS13505	C_RS13515	C_RS13520	C_RS13525	C_RS13530	C_RS13540	C_RS13545	C_RS13550	C_RS13565	C_RS13570	C_RS13575	C_RS13580	C_RS13585	C_RS13600	C_RS13605	C_RS13610	C_RS13615	C_RS13620	C_RS13625
C_RS14020	C_RS13685	C_RS13690	C_RS13695	C_RS13700	C_RS27295	C_RS13710	C_RS28730	C_RS13715	C_RS13805	C_RS13810	C_RS13825	C_RS28735	C_RS13840	C_RS13855	C_RS13875	C_RS13985	C_RS13990	C_RS13995	C_RS14005	C_RS14015
C_RS14205	C_RS14030	C_RS14035	C_RS14040	C_RS14045	C_RS14055	C_RS14060	C_RS14065	C_RS14070	C_RS14075	C_RS14080	C_RS14085	C_RS14090	C_RS14100	C_RS14105	C_RS14140	C_RS14155	C_RS14165	C_RS14170	C_RS14180	C_RS14195
C_RS14490	C_RS14335	C_RS14340	C_RS14345	C_RS14365	C_RS14375	C_RS14380	C_RS14410	C_RS14415	C_RS14420	C_RS14430	C_RS14435	C_RS14440	C_RS14445	C_RS14450	C_RS14460	C_RS14465	C_RS14470	C_RS14475	C_RS14480	C_RS14485
C_RS14720	C_RS14495	C_RS14500	C_RS14505	C_RS14510	C_RS14515	C_RS14520	C_RS14530	C_RS14635	C_RS14640	C_RS14650	C_RS14655	C_RS14660	C_RS14665	C_RS14670	C_RS14675	C_RS14680	C_RS14685	C_RS14690	C_RS14695	C_RS14700
C_RS14905	C_RS14725	C_RS14730	C_RS14735	C_RS14740	C_RS14745	C_RS14815	C_RS14820	C_RS14825	C_RS14830	C_RS14835	C_RS14840	C_RS14845	C_RS14865	C_RS14870	C_RS14875	C_RS14880	C_RS14885	C_RS14890	C_RS14895	C_RS14900
C_RS15635	C_RS14910	C_RS14915	C_RS14920	C_RS14930	C_RS14935	C_RS15405	C_RS15410	C_RS15415	C_RS15455	C_RS15470	C_RS15520	C_RS15570	C_RS15580	C_RS15590	C_RS15595	C_RS15600	C_RS15605	C_RS15615	C_RS15620	C_RS15625
C_RS15925	C_RS15640	C_RS15655	C_RS15730	C_RS15735	C_RS15740	C_RS15745	C_RS15750	C_RS15755	C_RS15760	C_RS15765	C_RS15770	C_RS15780	C_RS15785	C_RS15790	C_RS15795	C_RS15855	C_RS15880	C_RS15885	C_RS15890	C_RS15920
C_RS16210	C_RS15930	C_RS15935	C_RS15945	C_RS15955	C_RS15960	C_RS15965	C_RS15970	C_RS15975	C_RS15990	C_RS16005	C_RS16010	C_RS16020	C_RS16025	C_RS16030	C_RS16035	C_RS16040	C_RS16045	C_RS16190	C_RS16195	C_RS16200
C_RS16495	C_RS16225	C_RS16235	C_RS16240	C_RS16245	C_RS16250	C_RS16265	C_RS16270	C_RS16275	C_RS16280	C_RS16290	C_RS16300	C_RS16305	C_RS16310	C_RS16320	C_RS16330	C_RS16335	C_RS16350	C_RS16360	C_RS16365	C_RS16485
C_RS16620	C_RS16500	C_RS16505	C_RS16510	C_RS16515	C_RS16525	C_RS16530	C_RS16535	C_RS16540	C_RS16545	C_RS16550	C_RS16555	C_RS16560	C_RS16565	C_RS16570	C_RS16575	C_RS16580	C_RS16585	C_RS16590	C_RS16595	C_RS16605
C_RS16850	C_RS16625	C_RS16630	C_RS16635	C_RS16640	C_RS16645	C_RS16680	C_RS16715	C_RS16720	C_RS16750	C_RS16755	C_RS16760	C_RS16770	C_RS16775	C_RS16785	C_RS16795	C_RS16805	C_RS16815	C_RS16830	C_RS16840	C_RS16845
C_RS19565	C_RS16855	C_RS16860	C_RS16870	C_RS27500	C_RS16875	C_RS19475	C_RS19480	C_RS19485	C_RS19490	C_RS19495	C_RS19500	C_RS19505	C_RS19510	C_RS19515	C_RS19520	C_RS19525	C_RS19540	C_RS19545	C_RS19555	C_RS19560
C_RS19755	C_RS19570	C_RS19575	C_RS19585	C_RS19590	C_RS19600	C_RS19605	C_RS19610	C_RS19615	C_RS19620	C_RS19625	C_RS19675	C_RS19685	C_RS19695	C_RS19700	C_RS19705	C_RS19710	C_RS19730	C_RS19735	C_RS19740	C_RS19750
C_RS19995	C_RS19760	C_RS19770	C_RS19775	C_RS19790	C_RS19795	C_RS19800	C_RS19805	C_RS19820	C_RS19830	C_RS19835	C_RS19840	C_RS19850	C_RS19855	C_RS19895	C_RS19950	C_RS19955	C_RS19960	C_RS19965	C_RS19970	C_RS19980
C_RS20165	C_RS20000	C_RS20005	C_RS20010	C_RS20015	C_RS28835	C_RS20055	C_RS20060	C_RS20065	C_RS20070	C_RS20085	C_RS20090	C_RS20095	C_RS20100	C_RS20105	C_RS20110	C_RS20120	C_RS20125	C_RS20130	C_RS20145	C_RS20160
C_RS20405	C_RS20170	C_RS20175	C_RS20180	C_RS20185	C_RS20190	C_RS20195	C_RS20200	C_RS20205	C_RS20265	C_RS20270	C_RS20290	C_RS20295	C_RS20300	C_RS20305	C_RS20310	C_RS20315	C_RS20320	C_RS20330	C_RS20340	C_RS20400
C_RS20640	C_RS20410	C_RS20415	C_RS20460	C_RS20470	C_RS20475	C_RS20480	C_RS20490	C_RS20500	C_RS20505	C_RS20510	C_RS20515	C_RS20575	C_RS20580	C_RS20585	C_RS20590	C_RS20595	C_RS20600	C_RS20610	C_RS20625	C_RS20635
C_RS20950	C_RS20655	C_RS20660	C_RS20665	C_RS20670	C_RS20675	C_RS20685	C_RS20690	C_RS20860	C_RS20865	C_RS20870	C_RS20875	C_RS20885	C_RS20915	C_RS20920	C_RS20925	C_RS20930	C_RS20935	C_RS20940	C_RS28850	C_RS20945
C_RS21135	C_RS20965	C_RS20970	C_RS20975	C_RS20980	C_RS21035	C_RS21040	C_RS21050	C_RS21055	C_RS21060	C_RS21065	C_RS21075	C_RS21080	C_RS21085	C_RS21090	C_RS21100	C_RS21110	C_RS21115	C_RS21120	C_RS21125	C_RS21130
C_RS21795	C_RS21140	C_RS21145	C_RS21620	C_RS21625	C_RS21635	C_RS21640	C_RS21645	C_RS21650	C_RS21670	C_RS21675	C_RS21680	C_RS21685	C_RS21690	C_RS21715	C_RS21725	C_RS21730	C_RS21735	C_RS21740	C_RS21780	C_RS21785
C_RS21995	C_RS21800	C_RS21805	C_RS21815	C_RS21820	C_RS21830	C_RS21835	C_RS21920	C_RS21925	C_RS21930	C_RS21935	C_RS21940	C_RS21945	C_RS21950	C_RS21960	C_RS21965	C_RS21970	C_RS21975	C_RS21980	C_RS21985	C_RS21990
C_RS22210	C_RS22000	C_RS22010	C_RS22015	C_RS22125	C_RS22130	C_RS22135	C_RS22140	C_RS22145	C_RS22150	C_RS22155	C_RS22160	C_RS22165	C_RS22170	C_RS22175	C_RS27900	C_RS22180	C_RS22185	C_RS22190	C_RS22195	C_RS22205
C_RS22410	C_RS22215	C_RS22220	C_RS22240	C_RS22245	C_RS22250	C_RS22255	C_RS22260	C_RS22265	C_RS22300	C_RS22315	C_RS27905	C_RS22320	C_RS22325	C_RS22340	C_RS22345	C_RS22350	C_RS22355	C_RS22360	C_RS22365	C_RS22400
C_RS22745	C_RS22420	C_RS22425	C_RS22435	C_RS22450	C_RS22540	C_RS22545	C_RS22580	C_RS22585	C_RS22590	C_RS22595	C_RS22600	C_RS22605	C_RS22610	C_RS22615	C_RS22625	C_RS22630	C_RS22635	C_RS22725	C_RS22735	C_RS22740
C_RS23015	C_RS22750	C_RS22755	C_RS22760	C_RS22765	C_RS28905	C_RS22775	C_RS22845	C_RS22850	C_RS22885	C_RS22890	C_RS22895	C_RS22910	C_RS22960	C_RS22980	C_RS22985	C_RS22990	C_RS22995	C_RS23000	C_RS23005	C_RS23010
C_RS23650	C_RS23020	C_RS27940	C_RS23025	C_RS23030	C_RS23035	C_RS23040	C_RS23045	C_RS23050	C_RS23055	C_RS23060	C_RS23065	C_RS23070	C_RS23075	C_RS23080	C_RS23085	C_RS23100	C_RS23565	C_RS23585	C_RS23635	C_RS27975
C_RS23950	C_RS23700	C_RS23710	C_RS23725	C_RS23730	C_RS23735	C_RS23740	C_RS23745	C_RS23750	C_RS23755	C_RS23760	C_RS23770	C_RS23805	C_RS23830	C_RS23895	C_RS23900	C_RS23905	C_RS23915	C_RS23920	C_RS23925	C_RS23930
C_RS25125	C_RS23975	C_RS23980	C_RS23985	C_RS23995	C_RS24000	C_RS24180	C_RS24185	C_RS24215	C_RS24260	C_RS24655	C_RS24660	C_RS24675	C_RS24685	C_RS24690	C_RS24695	C_RS24700	C_RS25090	C_RS25100	C_RS25105	C_RS25120
C_RS25320	C_RS25130	C_RS25140	C_RS25150	C_RS25165	C_RS25170	C_RS25175	C_RS25180	C_RS25190	C_RS25195	C_RS25205	C_RS25215	C_RS25225	C_RS25230	C_RS25235	C_RS25240	C_RS25295	C_RS25300	C_RS25305	C_RS25310	C_RS25315
C_RS25935	C_RS25695	C_RS25700	C_RS25705	C_RS25715	C_RS25820	C_RS25825	C_RS25830	C_RS25835	C_RS25860	C_RS25865	C_RS25870	C_RS25885	C_RS25890	C_RS25895	C_RS25905	C_RS25910	C_RS25915	C_RS25920	C_RS25925	C_RS25930

Table 6.1. List of the 1092 *E. coli* core-genome target genes, which were included in the cgMLST typing scheme. These genes were compiled from 370 complete *E. coli* genomes on NCBI GenBank.

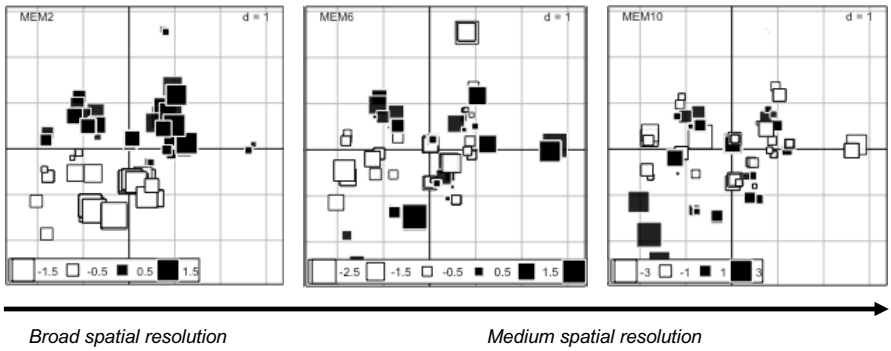


Figure 6.1. Significant distance-based Moran's eigenvector maps (dbMEM) eigenvectors from the model in eqn (3). Colour and size of squares on each plot indicate whether the spatial elements in each eigenvector are positively or negatively weighted, and taken together indicate the scale across which that eigenvector models spatial variation. Numbering of eigenvectors (2, 6 10, 13, 16) indicates the scale at which each eigenvector models spatial variation on a numerical scale between 1 and 17 (the total number of eigenvectors modelling positive spatial structure in the dataset).

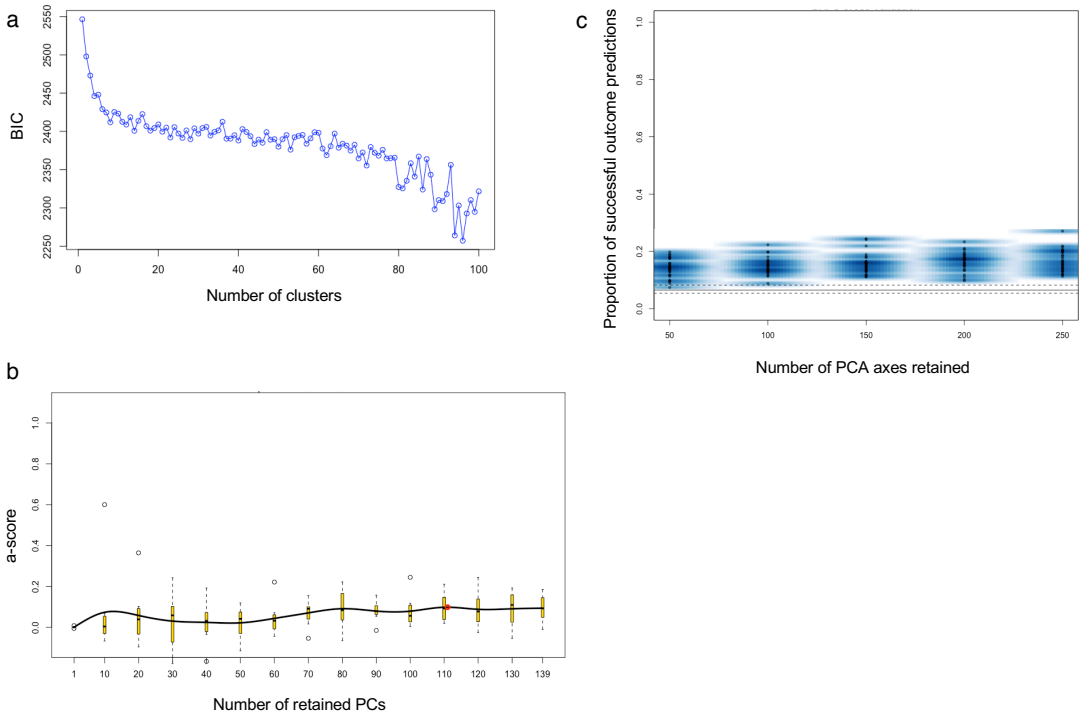


Figure 6.2. **a)** Plot showing use of Bayesian Information Criteria (BIC) to determine an appropriate number of clusters in the discriminant analysis of principal components (DAPC) performed on wildlife *E. coli* sequences. The BIC value continues to descend beyond 100 clusters, and as such, the estimate that $K = 6$ groups represent the best summary of the data is inconclusive. **b)** Optimisation α-score graph for the wildlife-only dataset, indicating that roughly 111 principal components (PCs) should be retained in DAPC, and **c)** cross-validation results for the combined dataset featuring all compartments. The highest proportion of successful outcomes for the simulation test occurs when roughly 200 PCs are retained.

	Pig	Rodent	Human	Bovid	Avian Seed	Bat Fruit/Nectar	Bat Invertebrate	Avian Omnivore	Chicken	Fowl	Small Ruminant	Outdoor Environment	Avian Invertebrate	Livestock Pens	Avian Scavenger	Avian Fruit/Nectar
Pig	0.000															
Rodent	0.033	0.000														
Human	0.008	0.021	0.000													
Bovid	0.024	0.012	0.018	0.000												
Avian Seed	0.016	0.009	0.009	0.005	0.000											
Bat Fruit/Nectar	0.076	0.051	0.032	0.062	0.043	0.000										
Bat Invertebrate	0.053	0.007	0.031	0.003	0.005	0.058	0.000									
Avian Omnivore	0.034	0.013	0.010	0.016	0.006	0.004	0.002	0.000								
Chicken	0.000	0.028	0.008	0.022	0.011	0.058	0.042	0.024	0.000							
Fowl	0.001	0.046	0.013	0.043	0.028	0.062	0.074	0.039	0.003	0.000						
Small Ruminant	0.052	0.017	0.038	0.016	0.018	0.073	-0.008	0.024	0.046	0.069	0.000					
Outdoor Environment	0.003	0.020	0.006	0.011	0.005	0.063	0.022	0.016	0.002	0.014	0.033	0.000				
Avian Invertebrate	0.008	0.021	0.008	0.016	0.005	0.019	0.026	0.002	0.010	0.017	0.031	0.013	0.000			
Livestock Pens	0.001	0.022	0.007	0.015	0.006	0.058	0.035	0.020	-0.001	0.005	0.038	0.001	0.008	0.000		
Avian Scavenger	0.000	0.033	-0.011	0.022	0.005	0.026	0.043	0.005	-0.004	-0.001	0.046	0.000	-0.009	-0.004	0.000	
Avian Fruit/Nectar	0.098	0.068	0.055	0.080	0.063	-0.011	0.064	0.014	0.082	0.091	0.086	0.078	0.043	0.081	0.051	0.000

Table 6.2. Pairwise Weir and Cockerham F_{ST} values, between wildlife functional groups, livestock taxa, humans and environmental groups. F_{ST} is measured on a scale of 0 to 1 (minimally and maximally divergent, respectively).

Chapter 7

General Discussion

7 General Discussion

Emerging diseases that spillover into new vertebrate hosts present a considerable economic and epidemiological burden on society. The process by which spillover occurs can be thought of at two scales; *i)* the ecological influence of landscape changes on the distribution and density of reservoir and target hosts (and thus the 'pathogen pressure' on target hosts), and *ii)* a series of epidemiological and behavioural limiting steps (or 'barriers') to transmission between reservoir and target hosts, and replication in and dissemination amongst target hosts (reviewed by Plowright *et al.* [132]). An expansive body of theory exists to describe processes acting at the former scale, but there is little empirical evidence that directly links changes in the function of abiotic and biotic systems to the structure of host communities, and dynamics of parasite communities living within them. In this thesis, I used data collected as part of a large field study in Nairobi to investigate how broad-scale ecological factors associated with rapid urban land-use change influence the ecology and epidemiology of multi-host communities. As a likely point of contact (and thus parasite transmission) between vertebrate wildlife, livestock, and humans, household 'interfaces' were chosen as sampling units. Being widely distributed across the urban environment, these interfaces capture much of the variation in urban land use, making them an ideal proxy through which to decompose the ecological processes that influence the form of their host and parasite communities. In the discussion that follows, I provide a brief summary of the main results of this thesis, before drawing on the results and discussion presented in earlier chapters to discuss the broader implications of these results to the field of community epidemiology, and practical implications for animal, public and ecosystem health with relation to urban development.

7.1 Thesis summary

The initial objective in this thesis was to use a biologically scaled approach to decompose the ecological and anthropogenic processes that influence assemblages of wildlife, livestock and human hosts at urban wildlife-livestock interfaces in Nairobi. By characterising broad-scale processes acting on household interfaces, and relating

them to variation in the diversity, density and assemblage of hosts, I demonstrated that opposing gradients of ecological and anthropogenic influence (e.g. increasing biotic habitat diversity, decreasing artificial land use and decreasing wealth) have measurable impacts upon the diversity, density and species assemblage of host compartments (wildlife, livestock and humans) (Chapter 4). On this basis, household interfaces could be characterised into one of four groups, which captured structural characteristics of their multi-host communities, and the ecological/anthropogenic processes that influence this community structure. Network analysis, examining the potential of different wildlife and livestock taxa to host and share zoonotic parasites, was used to relate zoonotic parasite sharing to the structural changes in host communities along an urban gradient.

In Chapters 5 and 6, I considered the epidemiological role of multi-host wildlife communities at urban interfaces. Utilising phenotypic and genetic characterisation of *E. coli*, the analyses in these chapters explored the epidemiological relationship between urban wildlife taxa occupying different functional niches, and more broadly between urban wildlife, livestock and humans. Chapter 5 focused on the epidemiology of clinically relevant antimicrobial resistance (AMR) in urban wildlife, exploring variation in carriage of AMR-*E. coli* between different wildlife taxa, and relating carriage in wildlife to other epidemiological compartments at interfaces. I demonstrated that urban wildlife are a net recipient of AMR from livestock and humans, with exposure being determined by feeding ecology. Exchange of AMR between synanthropic wildlife, livestock and humans provided empirical evidence for epidemiological connectivity between host compartments at household interfaces. In Chapter 6, high resolution genetic data was used to characterise *E. coli* populations, and infer bacterial transmission between wildlife taxa, and between wildlife, livestock, humans and the environment. Clear population differentiation was only present between wildlife taxa with distinct feeding ecology, and as such my results suggest that *E. coli* is circulated widely between host compartments at interfaces. Finally, by comparing variation in the diversity of mobile elements within the genome of wildlife *E. coli* isolates to the structural form of the host communities to which they belong, epidemiological processes were linked to their ecological and anthropogenic drivers (identified in Chapter 4). In doing so, I demonstrated that communities of microbial genes respond to changes in the richness and density of vertebrate host communities,

variation in which can provide a mechanism for spillover between epidemiologically connected wildlife, livestock and humans at interfaces.

7.2 Epidemiological consequences of structural change in host communities at urban interfaces

Previous studies have linked different elements of land-use change to infectious disease systems (e.g. cutaneous leishmaniasis, *Giardia sp. ,* leptospirosis, and *E. coli* [80,239–241]), but very few identify causal processes that link landscape-scale change and pathogen dynamics (with the notable exception of work conducted on Hendravirus in fruit bats, and Chytridiomycosis in amphibians [31,38,242] (reviewed by Gottdenker *et al.* [243])). By investigating such causal processes from a systems based perspective (that are not driven by interest in the epidemiology of a single pathogen) the approach used in this thesis is novel, with results that are broadly applicable to multi host-parasite communities, and not limited to the relationship between a single parasite and its host(s). The key findings presented in Chapters 5-6 are combined in Figure 7.1 to demonstrate the influence of abiotic factors deriving from urban land-use change on host and parasite biotic niches at wildlife-livestock-human interfaces. In this schematic, the epidemiological consequences of changes to the biotic niche occupied by communities of parasites (i.e. host community) are demonstrated by perturbations in the genetic and phenotypic profiles of *E. coli*. These results relate directly to the conceptual frameworks of urban disease emergence presented in Chapter 2 and Hassell *et al.* [9], and validate key elements of these frameworks. As such they could be used to guide further research into urban disease emergence, and inform surveillance efforts for zoonotic diseases in urban settings.

Most importantly, the results in this thesis provide empirical evidence that urban land-use change, through the influence of habitat alteration and human social factors on host assemblages at wildlife-livestock-interfaces, influences the genetic structure of parasite communities, and increases epidemiological connectivity between host compartments. Being dictated by a complex set of pathogen, host and site-specific factors (reviewed by Plowright *et al.* [132]), prediction of pathogen spillover between wildlife, livestock and humans in Nairobi is beyond the remit of this project. However,

by identifying the broader-scale factors that promote epidemiological connectivity within households (demonstrated here by exchange of AMR-*E. coli* genes and phenotypes between synanthropic wildlife and livestock), I have shown that changes in the structure of host communities on either side of an interface can fulfil two of the three epidemiological requirements for directly transmitted pathogens to spillover; sufficient reservoir-target contact, and an adequate prevalence of disease (i.e. AMR-*E. coli*) within the reservoir [44]. These results (in particular the categorisation of households into 'modes' as depicted in Figure 7.1) can be translated to help structure future research and guide surveillance efforts and evidence-based interventions to reduce the risk of pathogen transfer between host species. In order to refine the results of this project, and fully characterise the risk of pathogen spillover at urban wildlife-livestock-human interfaces, research focusing on identifying variation in interface-scale 'barriers' to transmission (i.e. wildlife abundance/density, immunological tolerance of host compartments, risk factors for human exposure to wildlife and livestock) and surveillance/characterisation of parasite communities within hosts (e.g. through metagenomic sequencing) is required. Structuring such research by the household modes identified in this study, would enable interface-scale risk factors to be linked to landscape-scale drivers.

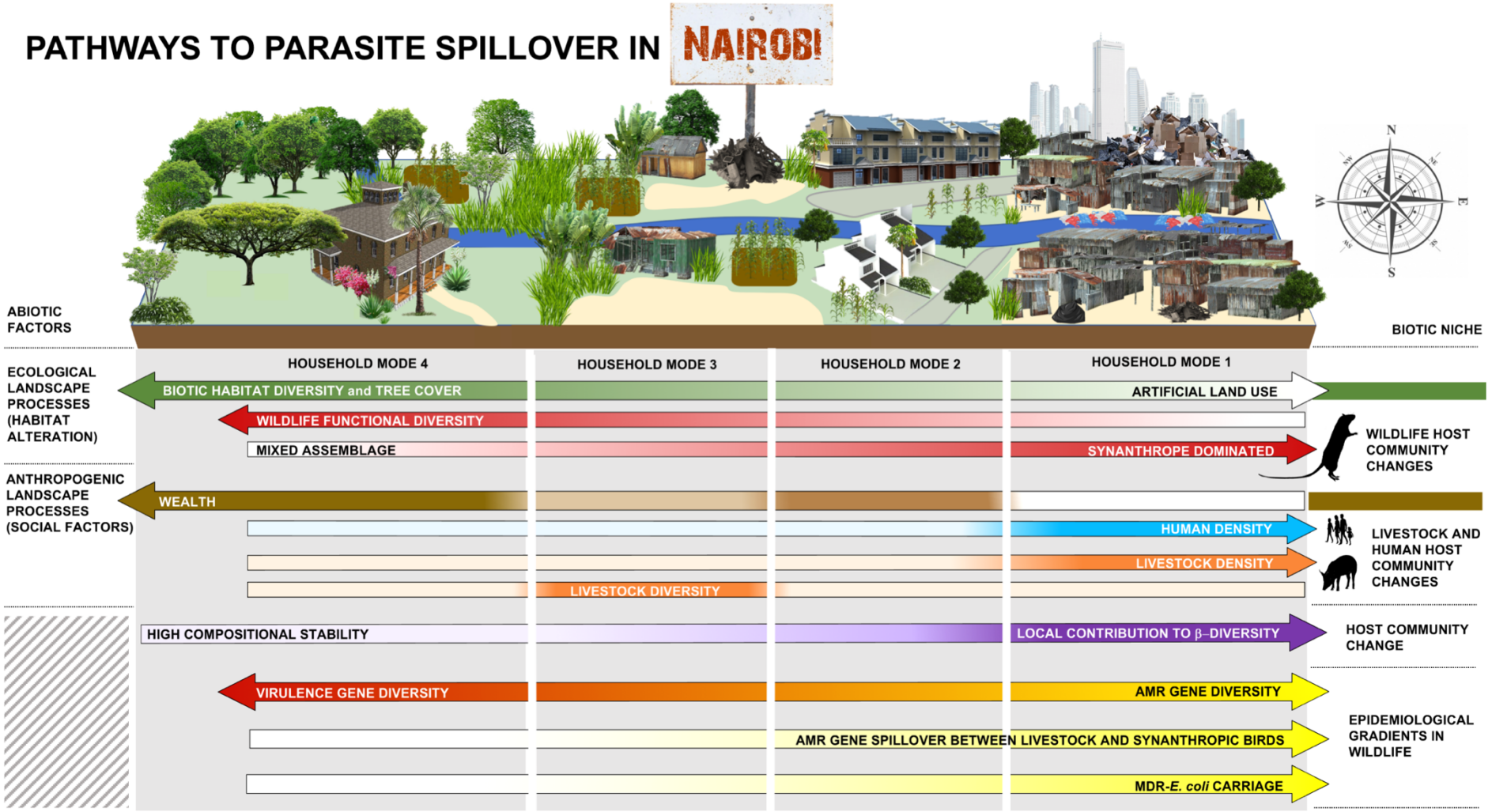


Figure 7.1. Diagrammatic representation of cascading biotic and abiotic influences on host community dynamics and subsequent epidemiological responses of *E. coli* in four categories ('modes') of household wildlife-livestock-human interface identified in Nairobi. This schematic brings together the main results from the thesis, to present a unified picture of the ecological and epidemiological effects of urban land-use change.

7.3 Implications for surveillance, public health and urban planning

Due to the expertise required and cost involved in conducting routine disease surveillance on wildlife, the prevalence of urban wildlife-borne zoonoses in developing countries is poorly described. Citizen science projects, aimed at gathering reports from the public to detect and report mortality in wildlife, are active in a number of European countries [244]. Such systems could offer a cost-effective model to conduct routine surveillance for wildlife-borne zoonoses in developing cities, and could be integrated with existing systems for reporting of zoonoses in livestock and humans. The cost effectiveness of such a surveillance program would be increased by focusing on species deemed to be more important hosts (i.e. the 'pathogen pool') and interfaces that are more likely points of contact between reservoir-target hosts. Based on the results from this project, household surveillance in Nairobi would be best focused on rodents, seed-eating and scavenging birds, and insectivorous bats, in low income, livestock keeping households, composed of high densities of humans, livestock and synanthropic wildlife (i.e. households belonging to mode 1 in Figure 7.1).

Given their epidemiological significance, interfaces also represent a critical point of control for the transmission of zoonoses, livestock disease, and pathogens that threaten vulnerable wildlife populations. To reduce the risk of disease transmission between wildlife, livestock and humans, interventions could be implemented at an interface, or policy level [245]. For example, in Chapter 4, resource provisioning (through livestock manure and human rubbish) was identified as a risk factor for exchange of AMR between synanthropic wildlife, and livestock and humans. An intervention aimed at educating people not to dispose of livestock and human waste within households could be targeted at residents of Nairobi, although, given that transfer of clinically relevant AMR is likely from livestock and humans to wildlife, the impact of such an intervention on human and livestock health would likely be minimal. The effects of resource provisioning on pathogen dynamics in wildlife is variable [246]. As such, without further characterisation of contact between wildlife, livestock and humans, and exchange of pathogens between them, it is not possible to translate the findings of this research into policy recommendations aimed at reducing

transmission of diseases that threaten wildlife, livestock and human health at interfaces in Nairobi.

The most consistent change in wildlife community structure across Nairobi and other urban areas is loss of species and functional diversity, and thus loss of host diversity [8]. In Nairobi, this occurs along a gradient of decreasing tree cover and biotic habitat diversity, and increasing artificial land use. As generally accepted in the scientific literature [157] (and demonstrated by proxy using virulence genes in this thesis), concurrent loss of host and parasite diversity reduces the 'pathogen pool' to which humans and livestock could potentially be exposed. However, biodiversity loss can influence the transmission dynamics of certain (predominantly vector borne) parasites, increasing human disease risk through a process known as the dilution effect [247]. Experimental models have demonstrated dilution effects for several widely distributed pathogens with rodent reservoirs, including Lyme disease, *Bartonella* and *Hantavirus* [248]. Given the shift towards synanthrope dominated communities as wildlife diversity decreases along an urban gradient in Nairobi, increases in the relative abundance of key reservoir hosts (such as rodents or bats) could increase parasite transmission, if vectors feed on a higher percentage of hosts that are capable of contracting, amplifying and transmitting the pathogen. Similarly, increasing abundance of competent wildlife hosts for zoonotic pathogens could result in amplification of pathogens [249]. Such compositional changes in small rodent communities, instigated by habitat conversion and loss of large mammals, have been linked to rodent-borne pathogen prevalence [250]. The influence of changes in wildlife community structure on disease prevalence have not been systematically explored along gradients of urbanisation, where the dilution effect or other changes in pathogen dynamics could be investigated either through experimental or observational studies. Such studies would be very important in informing urban planners of the epidemiological importance of maintaining vertebrate biodiversity in cities, and the mechanisms (such as reforestation) by which this can be achieved. More broadly, maintaining urban biodiversity in an ecologically sensitive manner could offer a suite of ecological services. Through pollination, seed-dispersal, buffering against the invasion of exotic wildlife species, improving atmospheric quality, and improving human quality of life, urban biodiversity can improve the health and livelihoods of people and wildlife living in urban environments (reviewed by Goddard *et al.* [251]).

7.4 Future Directions and Closing Remarks

Global databases of host-parasite associations are used extensively in the study of zoonotic pathogen emergence. Such studies tend to focus on describing broad characteristics of emerging zoonoses: host and pathogen traits that predict zoonotic carriage/potential, geographic distribution and host range, and anthropogenic and ecological drivers [60,125,135,252,253]. Whilst these studies can provide valuable insight into the epidemiology of zoonotic parasites, and generate data that will aid in guiding research efforts and pathogen surveillance at a global scale, models utilising data from carefully planned experimental studies that consider site-specific processes underlying zoonotic spillover are lacking, and are currently limited to in-depth epidemiological field studies of selected hosts and pathogens (e.g. [38]).

Given the current rate of urbanisation, and the potential for associated changes in societal structure, food systems and natural ecosystems to expose human and animal populations to novel pathogens, I recommend an interdisciplinary approach to studying urban human-wildlife-livestock interfaces in Chapter 2, with the following aims: *i)* "establish characterisations for potential 'high risk' interfaces that exist along gradients of urbanisation, and identify processes that have led to their formation", *ii)* "describe biological organisation and community ecology at these interfaces, conduct surveillance for priority zoonotic pathogens (i.e. those with 'emergent potential') across host taxa, and study the evolutionary processes underlying cross-species transmission where it is detected", *iii)* "at interfaces where transmission risks are identified, develop appropriate interventions that can be used to reduce risk of transmission". The methods and analysis contained within this thesis successfully address the first, and parts of the second, of these aims for urban households, and could easily be extended to other urban settings, and urban wildlife-livestock-human interfaces such as value chain nodes. However, although *E. coli* phenotypes and genotypes are useful markers of epidemiological connectivity, they provide little information on the structure of parasite communities within hosts, how these communities respond to changes at higher trophic levels, and how this influences parasite sharing between host species. Such research questions represent the next step towards fully characterising the risk of disease transmission at interfaces, and could be addressed by adopting a community epidemiology approach [254]. Recent advances

in metagenomics permit sequencing of bacterial and viral microbiomes, which, when analysed with cutting edge population genetic and/or phylogenetic tools, could be used to explore variation in the structure of parasite communities within and between host populations. Combined with ecological and anthropological characterisation of interfaces (and ideally indicators for interface-scale barriers to transmission described by Plowright *et al.* [132]), such research would bring us a step closer to understanding the influence of urbanisation on disease transmission. This would provide more reliable data with which to define surveillance protocols and interventions to mitigate the risk of transmission. Decomposing the intricate nature of urban processes to deliver evidence-based policy that improves wildlife, livestock and human health, will require a transdisciplinary effort, spanning ecology, epidemiology, anthropology, sociology and urban planning.

Chapter 8

Bibliography

1 Mather, A.E. *et al.* (2012) An ecological approach to assessing the epidemiology of antimicrobial resistance in animal and human populations. *Proc. R. Soc. B.* 279, 1630–9

2 Alirol, E. *et al.* (2011) Urbanisation and infectious diseases in a globalised world. *Lancet Infect. Dis.* 11, 131–41

3 UN Department of Economic and Social Affairs Population Division (2014) World urbanization prospects: the 2014 revision http://esa.un.org/unpd/wup/Highlights/WUP2014-Highlights.pdf

4 Taylor, L.H. *et al.* (2001) Risk factors for human disease emergence. *Philos. Trans. R. Soc. B.* 356, 983–9

5 Woolhouse, M.E.J. and Gowtage-Sequeria, S. (2005) Host range and emerging and reemerging pathogens. *Emerg. Infect. Dis.* 11, 1842–7

6 Cole, D. *et al.* (2008) Healthy city harvests: generating evidence to guide policy on urban agriculture, International Potato Centre, Lima

7 Mwangi, A.M. (1995) The Role of Urban Agriculture for Food Security in Low Income Areas in Nairobi, FNSP report, African Studies Centre

8 Bradley, C.A. and Altizer, S. (2007) Urbanization and the ecology of wildlife diseases. *Trends Ecol. Evol.* 22, 95–102

9 Hassell, J.M. *et al.* (2017) Urbanization and Disease Emergence: Dynamics at the Wildlife–Livestock–Human Interface. *Trends Ecol. Evol.* 32, 55–67

10 Smolinsky, M.S. *et al.* (2003) Microbial threats to health: Emergence, detection, and response (Smolinski M.S., Hamburg M.A. and Lederberg J. eds), National Academies Press

11 Jones, K.E. *et al.* (2008) Global trends in emerging infectious diseases. *Nature* 451, 990–3

12 Karesh, W.B. *et al.* (2012) Ecology of zoonoses: natural and unnatural histories. *Lancet* 380, 1936–45

13 McMichael, a J. (2004) Environmental and social influences on emerging infectious diseases: past, present and future. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 359, 1049–58

14 Daszak, P. (2000) Emerging Infectious Diseases of Wildlife- Threats to Biodiversity and Human Health. *Science* 287, 443–449

15 Perry, B.D. *et al.* (2011) Livestock and Global Change Special Feature: Current drivers and future directions of global livestock disease dynamics. *Proc. Natl. Acad. Sci. USA* 110, 20871–20877

16 Jones, B. a *et al.* (2013) Zoonosis emergence linked to agricultural intensification and environmental change. *Proc. Natl. Acad. Sci. USA* 110, 8399–404

17 Patz, J. a. *et al.* (2004) Unhealthy Landscapes: Policy Recommendations on Land Use Change and Infectious Disease Emergence. *Environ. Health Perspect.* 112, 1092–1098

18 McFarlane, R. a *et al.* (2013) Land-use change and emerging infectious disease on an island continent. *Int. J. Environ. Res. Public Health* 10, 2699–719

19 Knobler, S. *et al.* (2006) The Impact of Globalization on Infectious Disease Emergence and Control: Exploring the consequences and opportunities. *Proceedings of the Forum on Microbial Threats* (Knobler s., Mahmoud A., Lemon S. and Pray L. eds), pp 21-48, Institute of Medicine, Washington DC

20 UN Department of Economic and Social Affairs Population Division. World urbanization prospects: the 2014 revision (2014) http://esa.un.org/unpd/wup/Highlights/WUP2014-Highlights.pdf

21 Alberti, M. (2005) The Effects of Urban Patterns on Ecosystem Function. *Int. Reg. Sci. Rev.* 28, 168–192

22 Pickett, S.T.A. *et al.* (2011) Urban ecological systems: scientific foundations and a decade of progress. *J. Environ. Manage.* 92, 331–62

23 Lowry, H. *et al.* (2012) Behavioural responses of wildlife to urban environments. *Biol. Rev. Camb. Philos. Soc.* 88, 537–49

24 Deplazes, P. *et al.* (2004) Wilderness in the city: The urbanization of Echinococcus multilocularis. *Trends Parasitol*. 20, 77–84

25 Wright, A.N. and Gompper, M.E. (2005) Altered parasite assemblages in raccoons in response to manipulated resource availability. *Oecologia* 144, 148–156

26 Guan, P. *et al.* (2009) Investigating the effects of climatic variables and reservoir on the incidence of hemorrhagic fever with renal syndrome in Huludao City, China: a 17-year data analysis based on structure equation model. *BMC Infect. Dis.* 9, 109

27 Evangelista, K. and Coburn, J. (2010) Leptospira as an emerging pathogen: A review of its biology, pathogenesis, and host immune responses. *Future Microbiol.* 5, 1413–1425

28 Halliday, J.E.B. *et al.* (2013) Urban leptospirosis in Africa: a cross-sectional survey of Leptospira infection in rodents in the Kibera urban settlement, Nairobi, Kenya. *Am. J. Trop. Med. Hyg.* 89, 1095–102

29 Himsworth, C.G. *et al.* (2013) Rats, cities, people, and pathogens: a systematic review and narrative synthesis of literature regarding the ecology of rat-associated zoonoses in urban centers. *Vector Borne Zoonotic Dis.* 13, 349–59

30 Pulliam, J.R.C. *et al.* (2012) Agricultural intensification, priming for persistence and the emergence of Nipah virus: a lethal bat-borne zoonosis. *J. R. Soc. Interface* 9, 89–101

31 Plowright, R.K. *et al.* (2011) Urban habituation, ecological connectivity and epidemic dampening: the emergence of Hendra virus from flying foxes (Pteropus spp.). *Proc. R. Soc. B* 278, 3703–3712

32 Hall, S.R. *et al.* (2005) Selective predation and productivity jointly drive complex behavior in host-parasite systems. *Am. Nat.* 165, 70–81

33 Telfer, S. *et al.* (2010) Species interactions in a parasite community drive infection risk in a wildlife population. *Science* 330, 243–6

34 Tomczuk, K. *et al.* (2015) Analysis of intrinsic and extrinsic factors influencing the dynamics of bovine Eimeria spp. from central-eastern Poland. *Vet. Parasitol.* 214, 22–28

35 Begon, M. *et al.* (2009) Seasonal host dynamics drive the timing of recurrent epidemics in a wildlife population. *Proc. R. Soc. B* 276, 1603–1610

36 Keesing, F. *et al.* (2006) Effects of species diversity on disease risk. *Ecol. Lett.* 9, 485–98

37 Daszak, P. *et al.* (2006) The emergence of Nipah and Hendra virus: pathogen dynamics across a wildlife-livestock-human continuum. In *Disease ecology: community structure and pathogen dynamics* (Collinge S. and Ray C. eds), pp. 186–201, Oxford University Press

38 Plowright, R.K. *et al.* (2015) Ecological dynamics of emerging bat virus spillover. *Proc. R. Soc. B* 282, 20142124

39 Hayman, D.T.S. *et al.* (2013) Ecology of zoonotic infectious diseases in bats: current knowledge and future directions. *Zoonoses Public Health* 60, 2–21

40 Viana, M. *et al.* (2014) Assembling evidence for identifying reservoirs of infection. *Trends Ecol. Evol.* 29, 270–279

41 Caron, A. *et al.* (2015) Bridge hosts, a missing link for disease ecology in multi-host systems. *Vet. Res.* 46, 83

42 Begon, M. *et al.* (2002) A clarification of transmission terms in host-microparasite models: numbers, densities and areas. *Epidemiol. Infect.* 129, 147–53

43 Davis, S. and Calvet, E. (2005) Fluctuating rodent populations and risk to humans from rodent-borne zoonoses. *Vector-Borne Zoonotic Dis.* 5, 305–314

44 Lloyd-Smith, J.O. *et al.* (2009) Epidemic dynamics at the human-animal interface. *Science* 326, 1362–7

45 Gardner-Santana, L.C. *et al.* (2009) Commensal ecology, urban landscapes, and their influence on the genetic characteristics of city-dwelling Norway rats (Rattus norvegicus). *Mol. Ecol.* 18, 2766–78

46 Murray, K. and Daszak, P. (2013) Human ecology in pathogenic landscapes: two hypotheses on how land use change drives viral emergence. *Curr. Opin. Virol.* 3, 79–83

47 Lloyd-Smith, J.O. *et al.* (2014) Nine challenges in modelling the emergence of novel pathogens. *Epidemics* 10, 35-39

48 Faeth, S.H. *et al.* (2005) Trophic Dynamics in Urban Communities. *Bioscience* 55, 399

49 Galbraith, J. a *et al.* (2015) Supplementary feeding restructures urban bird communities. *Proc. Natl. Acad. Sci. USA* 112, e2648-2657

50 McKinney, M.L. (2002) Urbanization, Biodiversity, and Conservation. *Bioscience* 52, 883

51 Evans, K.L. *et al.* (2009) Habitat influences on urban avian assemblages. *Ibis* 151, 19–39

52 Hudson, P.J. *et al.* (2006) Is a healthy ecosystem one that is rich in parasites? *Trends Ecol. Evol.* 21, 381–385

53 Wood, C.L. *et al.* (2014) Does biodiversity protect humans against infectious disease? *Ecology* 95, 817–832

54 Chaisiri, K. *et al.* (2010) Human-dominated habitats and helminth parasitism in Southeast Asian murids. *Parasitol. Res.* 107, 931–937

55 Lawson, B. *et al.* (2010) Epidemiology of salmonellosis in garden birds in England and Wales, 1993 to 2003. *Ecohealth* 7, 294–306

56 Hamer, S.A. *et al.* (2012) Wild birds as sentinels for multiple zoonotic pathogens along an urban to rural gradient in greater Chicago, Illinois. *Zoonoses Public Health* 59, 355–64

57 Ezenwa, V.O. *et al.* (2006) Avian diversity and West Nile virus: testing associations between biodiversity and infectious disease risk. *Proc. R. Soc. B* 273, 109–117

58 Palacios, G. *et al.* (2011) Human metapneumovirus infection in wild mountain gorillas, Rwanda. *Emerg. Infect. Dis.* 17, 711–713

59 Franklinos, L.H. V *et al.* (2015) Streptococcus pyogenes Infection in a Free-Living European Hedgehog (Erinaceus europaeus). *Ecohealth* 12, 689–692

60 Kreuder Johnson, C. *et al.* (2015) Spillover and pandemic properties of zoonotic viruses with high host plasticity. *Sci. Rep.* 5, 14830

61 Wood, J.L.N. *et al.* (2012) A framework for the study of zoonotic disease emergence and its drivers: spillover of bat pathogens as a case study. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 367, 2881–92

62 Estrada-Peña, A. *et al.* (2014) Effects of environmental change on zoonotic disease risk: an ecological primer. *Trends Parasitol.* 30, 205–14

63 Meentemeyer, R.K. *et al.* (2012) Landscape epidemiology of emerging infectious diseases in natural and human-altered ecosystems. *Annu. Rev. Phytopathol.* 50, 379–402

64 Olson, D.H. *et al.* (2013) Mapping the Global Emergence of Batrachochytrium dendrobatidis, the Amphibian Chytrid Fungus. *PLoS One* 8 doi:10.1371/journal.pone.0056802

65 Wellington, E.M.H. *et al.* (2013) The role of the natural environment in the emergence of antibiotic resistance in gram-negative bacteria. *Lancet Infect. Dis.* 13, 155–65

66 Dobson, A. (2004) Population dynamics of pathogens with multiple host species. *Am. Nat.* 164, 64–S78

67 Pedersen, A.B. and Davies, T.J. (2009) Cross-species pathogen transmission and disease emergence in primates. *Ecohealth* 6, 496–508

68 Poulin, R. (2004) Macroecological patterns of species richness in parasite assemblages. *Basic Appl. Ecol.* 5, 423–434

69 Reperant, L. a (2010) Applying the theory of island biogeography to emerging pathogens: toward predicting the sources of future emerging zoonotic and vector-borne diseases. *Vector Borne Zoonotic Dis.* 10, 105–110

70 Pepin, K.M. *et al.* (2010) Identifying genetic markers of adaptation for surveillance of viral host jumps. *Nat. Rev. Microbiol.* 8, 802–813

71 Woolhouse, M.E.J. et al. (2005) Emerging pathogens: The epidemiology and evolution of species jumps. Trends Ecol. Evol. 20, 238–244

72 Longdon, B. *et al.* (2011) Host phylogeny determines viral persistence and replication in novel hosts. *PLoS Pathog.* 79, e1002260

73 Benmayor, R. *et al.* (2009) Host Mixing and Disease Emergence. *Curr. Biol.* 19, 764–767

74 Gandon, S. (2004) Evolution of multihost parasites. *Evolution* 58, 455–469

75 Ellis, E.C. and Ramankutty, N. (2008) Putting people in the map: anthropogenic biomes of the world. *Front. Ecol. Environ.* 6, 439–447

76 Reisen, W.K. (2010) Landscape epidemiology of vector-borne diseases. *Annu. Rev. Entomol.* 55, 461–83

77 Wilcove, D.S. *et al.* (1986) Habitat Fragmentation in the temperate zone. In *Conservation Biology: The Science of Scarcity and Diversity* (Soulé, E. ed), pp. 237–256, Sinaure Associates

78 Despommier, D. *et al.* (2006) The Role of Ecotones in Emerging Infectious Diseases. *Ecohealth* 3, 281–289

79 Brown, H.E. *et al.* (2008) Remotely-sensed vegetation indices identify mosquito clusters of West Nile virus vectors in an urban landscape in the northeastern United States. *Vector Borne Zoonotic Dis.* 8, 197–206

80 Goldberg, T. *et al.* (2008) Forest fragmentation as cause of bacterial transmission among nonhuman primates, humans, and livestock, Uganda. *Emerg Infect. Dis.* 14, 1375–1382

81 Eisen, R.J. *et al.* (2007) A spatial model of shared risk for plague and hantavirus pulmonary syndrome in the southwestern United States. *Am. J. Trop. Med. Hyg.* 77, 999–1004

82 Hahn, M.B. *et al.* (2014) The role of landscape composition and configuration on Pteropus giganteus roosting ecology and Nipah virus spillover risk in Bangladesh. *Am. J. Trop. Med. Hyg.* 90, 247–255

83 Jobbins, S.E. *et al.* (2014) Leptospira interrogans at the human-wildlife interface in northern Botswana: a newly identified public health threat. *Zoonoses Public Health* 61, 113–23

84 Lawson, B. *et al.* (2014) Epidemiological evidence that garden birds are a source of human salmonellosis in England and Wales. *PLoS One* doi:10.1371/journal.pone.0088968

85 Swirski, A.L. *et al.* (2014) Spatial Epidemiology of Escherichia coli O157:H7 in Dairy Cattle in Relation to Night Roosts Of Sturnus vulgaris (European Starling) in Ohio, USA (2007-2009). *Zoonoses Public Health* 61, 427–35

86 Becker, D.J. *et al.* (2015) Linking anthropogenic resources to wildlife-pathogen dynamics: A review and meta-analysis. *Ecol. Lett.* 18, 483–495

87 Sims, L.D. *et al.* (2005) Origin and evolution of highly pathogenic H5N1 avian influenza in Asia. *Vet. Rec.* 157, 159–64

88 Meng, X.J. *et al.* (2009) Wild boars as sources for infectious diseases in livestock and humans. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 364, 2697–2707

89 Murray, M. *et al.* (2015) Poor health is associated with use of anthropogenic resources in an urban carnivore. *Proc. R. Soc. B* 282, 20150009

90 Becker, D. and Hall, R. (2014) Too much of a good thing : resource provisioning alters infectious disease dynamics in wildlife. *Biol. Lett.* 10, 20140309

91 Padgett, D.A. and Glaser, R. (2003) How stress influences the immune response. *Trends Immunol*. 24, 444–448

92 Ledant, M. (2011) Socio-Economical and Infrastrucural Mapping and Analysis of Nairobi: Technical Report, IFRA/GWOPA/UN-HABITAT, Nairobi, Kenya

93 Berghaus, R.D. et al. (2013) Enumeration of Salmonella and Campylobacter spp. in environmental farm samples and processing plant carcass rinses from commercial broiler chicken flocks. *Appl. Environ. Microbiol.* 79, 4106–14

94 Sbodio, A. et al. (2013) Modified Moore swab optimization and validation in capturing *E. coli* O157:H7 and *Salmonella enterica* in large volume field samples of irrigation water. *Food Res. Int.* 51, 654–662

95 Bettridge, J.M. et al. (2017) Soup-E.coli-field logistics-epi-and-genomics: A sampling strategy to capture bacterial diversity in a changing urban

environment, in Society for Veterinary Epidemiology and Preventative Medicine, Inverness, March 29-31

96 Johnson, K. and Oscar Rutstein, S. (2004) The DHS wealth index. DHS comparative reports no. 6. Calverton: ORC Macro. 2004

97 Wilman, H. et al. (2014) EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals. *Ecology* 95, 2027

98 Zuur, A.F. et al. (2010) A protocol for data exploration to avoid common statistical problems. *Methods Ecol. Evol.* 1, 3–14

99 Legendre, P. and Gallagher, E.D. (2001) Ecologically meaningful transformations for ordination of species data. *Oecologia* 129, 271–280

100 Borcard, D. et al. (2011) Numerical Ecology with R, Use R! series, Springer Science, New York

101 Legendre, P. and Andersson, M.J. (1999) Distance-based redundancy analysis: Testing multispecies responses in multifactorial ecological experiments. *Ecol. Monogr.* 69, 1–24

102 Borcard, D. et al. (1992) Partialling out the spatial component of ecological variation. *Ecology* 73, 1045–1055

103 Blanchet, G. et al. (2008) Forward selection of spatial explanatory variables. *Ecology* 89, 2623–2632

104 Okansen, J. et al. (2017) Community Ecology Package, R Foundation for Statistical Computing

105 Zuur, A.F. et al. (2007) Analysing Ecological Data, Springer, New York

106 Akaike, H. (1974) A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* 19, 716–723

107 Bates, D. et al. (2017) Linear Mixed-Effects Models using "Eigen" and S4, R Foundation for Statistical Computing

108 Skaug, H. et al. (2015) glmmADMB: Generalized Linear Mixed Models using AD Model Builder, R Foundation for Statistical Computing

109 Wood, S.N. (2011) Fast stable REML and ML estimation of semiparametric GLMs. *J. R. Stat. Soc. Ser. B.* 73, 3–36

110 Legendre, P. and Gauthier, O. (2014) Statistical methods for temporal and space-time analysis of community composition data. *Proc. R. Soc. B.* 281, 20132728

111 Borcard, D. et al. (2004) Dissecting the spatial structure of ecological data at multiple scales. *Ecology* 85, 1826–1832

112 Dray, S. et al. (2017) Adespatial: multivariate multiscale spatial analysis, R Foundation for Statistical Computing

113 Bivand, R. et al. (2014) rgdal: Bindings for the geospatial data abstraction library, R Foundation for Statistical Computing

114 Whittaker, R.H. (1960) Vegetation of the Siskiyou Mountains, Oregon and California. *Ecol. Monogr.* 30, 279–338

115 Legendre, P. and De Cáceres, M. (2013) Beta diversity as the variance of community data: Dissimilarity coefficients and partitioning. *Ecol. Lett.* 16, 951–963

116 Bloom, D.E. (2011) 7 billion and counting. *Science* 333, 562–569

117 Werner, P. (2011) The ecology of urban areas and their functions for species diversity. *Landsc. Ecol. Eng.* 7, 231–240

118 Kock, R. (2014) Drivers of disease emergence and spread: Is wildlife to blame? *Onderstepoort J Vet Res* 81, 4–7

119 Groffman, P.M. *et al.* (2014) Ecological homogenization of urban USA. *Front. Ecol. Environ.* 12, 74–81

120 Horton, R.A. *et al.* (2013) Wild birds carry similar Salmonella enterica serovar Typhimurium strains to those found in domestic animals and livestock. *Res. Vet. Sci.* 95, 45–48

121 Arnold, K.E. *et al.* (2016) 'Disperse abroad in the land': the role of wildlife in the dissemination of antimicrobial resistance. *Biol. Lett.* 12, 20160137

122 Cleaveland, S. *et al.* (2001) Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence. *Philos. Trans. R. Soc. B.* 356, 991–9

123 Gortazar, C. *et al.* (2016) Infections shared with wildlife: an updated perspective. *Eur J Wildlife Res.* 62, 511–525

124 McFarlane, R. *et al.* (2012) Synanthropy of wild mammals as a determinant of emerging infectious diseases in the Asian-Australasian region. *Ecohealth* 9, 24–35

125 Han, B.A. *et al.* (2015) Rodent reservoirs of future zoonotic diseases. *Proc. Natl. Acad. Sci. U. S. A.* 112, 7039–44

126 Soulsbury, C.D. and White, P.C.L. (2015) Human–wildlife interactions in urban areas: a review of conflicts, benefits and opportunities. *Wildl. Res.* 42, 3330

127 Satterthwaite, D. *et al.* (2010) Urbanization and its implications for food and farming. *Philos. Trans. R. Soc. B.* 365, 1–228

128 Thys, E. *et al.* (2005) Socio-economic determinants of urban household livestock keeping in semi-arid Western Africa. *J. Arid Environ.* 63, 475–496

129 Schiere, H. and van der Hoek, R. (2001) Livestock keeping in urban areas: A review of traditional technologies based on literature and field experiences. FAO Animal Production and Health Paper 151, FAO, Rome, Italy. FAO: http://www.fao.org/DOCREP/004/Y0500E/y0500e00.htm#toc

130 Conan, A. *et al.* (2012) Biosecurity measures for backyard poultry in developing countries: a systematic review. *BMC Vet. Res.* 8, 240

131 Wiens, J.A. *et al.* (2009) Niches, models, and climate change: Assessing the assumptions and uncertainties. *Proc. Natl. Acad. Sci. U S A* 106, 19729–19736

132 Plowright, R.K. *et al.* (2017) Pathways to zoonotic spillover. *Nat Rev Microbiol*, 15, 502–510

133 World Health Organization (2006) WHO child growth standards: length/height for age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age, methods and development. World Health Organization

134 Wardeh, M. *et al.* (2015) Database of host-pathogen and related species interactions, and their global distribution. *Sci. Data* 2, 150049

135 Olival, K.J. *et al.* (2017) Host and viral traits predict zoonotic spillover from mammals. *Nature* 546, 646–650

136 R Core Team (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

137 Weber, M. *et al.* (2011) Practical application of self-organizing maps to interrelate biodiversity and functional data in NGS-based metagenomics. *ISME J.* 5, 918–928

138 Chon, T.S. (2011) Self-Organizing Maps applied to ecological sciences. *Ecol. Inform.* 6, 50–61

139 Wehrens, R. and Buydens, L.M.C. (2007) Self- and super-organizing maps in R: The kohonen package. *J. Stat. Softw.* 21, 1–19

140 Whitaker, D. and Christman, M. (2014) clustsig: Significant Cluster Analysis. R Foundation for Statistical Computing

141 Hamel, L. and Brown, C.W. (2012) Bayesian probability approach to feature significance for infrared spectra of bacteria. *Appl. Spectrosc.* 66, 48–59

142 Chytrý, M. *et al.* (2002) Determination of diagnostic species with statistical fidelity measures. *J. Veg. Sci.* 13, 79–90

143 Novotny, V. *et al.* (2007) Low beta diversity of herbivorous insects in tropical forests. *Nature* 448, 692–695

144 Legendre, P. (2014) Interpreting the replacement and richness difference components of beta diversity. *Glob. Ecol. Biogeogr.* 23, 1324–1334

145 Blanchet, F.G. *et al.* (2008) Forward selection of explanatory variables. *Ecology* 89, 2623–2632

146 Godfrey, S.S. (2013) Networks and the ecology of parasite transmission: A framework for wildlife parasitology. *Int. J. Parasitol. Parasites Wildl.* 2, 235–245

147 Pilosof, S. *et al.* (2015) Potential Parasite Transmission in Multi-Host Networks Based on Parasite Sharing. *PLoS One* 10, e0117909

148 Rivera-Hutinel, A. *et al.* (2012) Effects of sampling completeness on the structure of plant – pollinator networks. *Ecology* 93, 1593–1603

149 Vavrek, M. J. (2011). Fossil: palaeoecological and palaeogeographical analysis tools. *Palaeontologia Electronica* 14, 16

150 Bastian, M. *et al.* (2009) Gephi: An Open Source Software for Exploring and Manipulating Networks. *Third Int. AAAI Conf. Weblogs Soc. Media* DOI: 10.1136/qshc.2004.010033

151 Smoot, M.E. *et al.* (2011) Cytoscape 2.8: New features for data integration and network visualization. *Bioinformatics* 27, 431–432

152 McIntyre, K.M. *et al.* (2014) A quantitative prioritisation of human and domestic animal pathogens in Europe. *PLoS One* 9, 1–9

153 Barwell, L.J. *et al.* (2015) Measuring beta diversity with species abundance data. *J. Anim. Ecol.* 84, 1112–1122

154 McKinney, M.L. (2006) Urbanization as a major cause of biotic homogenization. *Biol. Conserv.* 127, 247–260

155 Hooper, D.U. *et al.* (2012) A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature* 486, 105–108

156 Roche, B. *et al.* (2012) Linking community and disease ecology: the impact of biodiversity on pathogen transmission. *Philos. Trans. R. Soc. B.* 367, 2807–2813

157 Ostfeld, R.S. and Keesing, F. (2017) Is biodiversity bad for your health? *Ecosphere* 8, 1676-1673

158 O'Neill, J. (2016) Tackling drug-resistant infections globally: final report and recommendations. Review on Antimicrobial Resistance, London, UK

159 Woolhouse, M. *et al.* (2015) Antimicrobial resistance in humans, livestock and the wider environment. *Philos. Trans. R. Soc. B.* 370, 20140083

160 Robinson, T.P. *et al.* (2016) Antibiotic resistance is the quintessential One Health issue. *Trans. R. Soc. Trop. Med. Hyg.* 7, 377-380

161 Holmes, A.H. *et al.* (2016) Understanding the mechanisms and drivers of antimicrobial resistance. *Lancet* 387, 176–187

162 Allen, H.K. *et al.* (2010) Call of the wild: antibiotic resistance genes in natural environments. *Nat. Rev. Microbiol.* 8, 251–9

163 Ayukekbong, J.A. *et al.* (2017) The threat of antimicrobial resistance in developing countries: causes and control strategies. *Antimicrob. Resist. Infect. Control* 6, 47

164 Lim, C. *et al.* (2016) Epidemiology and burden of multidrug-resistant bacterial infection in a developing country. *Elife* 5, 1–18

165 Mitema, E. *et al.* (2001) An assessment of antimicrobial consumption in food producing animals in Kenya. *J. Vet. Pharmacol. Ther.* 24, 385–390

166 Blaser, M.J. *et al.* (2016) Toward a Predictive Understanding of Earth's Microbiomes to Address 21st Century Challenges. *Am. Soc. Microbiol.* 7, 1–16

167 Gaze, W.H. *et al.* (2013) Influence of Humans on Evolution and Mobilization of Environmental Antibiotic Resistome. *Emerg. Infect. Dis. J.* 19, e120871

168 Harvey, C.F. *et al.* (2002) Arsenic Mobility and Groundwater Extraction in Bangladesh. *Science* 298, 1602–1606

169 Cabello, F.C. (2006) Heavy use of prophylactic antibiotics in aquaculture: a growing problem for human and animal health and for the environment. *Appl. Environ. Microbiol.* 8, 137–1144

170 Lupo, A. *et al.* (2012) Origin and evolution of antibiotic resistance: The common mechanisms of emergence and spread in water bodies. *Front. Microbiol.* 3, 1–13

171 Vittecoq, M. *et al.* (2016) Antimicrobial resistance in wildlife. *J. Appl. Ecol.* 53, 519–529

172 Murray, C.G. and Hamilton, A.J. (2010) Perspectives on wastewater treatment wetlands and waterbird conservation. *J. Appl. Ecol.* 47, 976–985

173 Furness, L.E. *et al.* (2017) Wild small mammals as sentinels for the environmental transmission of antimicrobial resistance. *Environ. Res.* 154, 28–34

174 Gilliver, M.A. *et al.* (1999) Antibiotic resistance found in wild rodents. *Nature* 401, 233–234

175 Smith, S. *et al.* (2014) Antimicrobial resistant bacteria in wild mammals and birds: a coincidence or cause for concern? *Ir. Vet. J.* 67, 88

176 Alarcon, P. *et al.* (2017) Urban Livestock Keeping in the City of Nairobi: Diversity of Production Systems, Supply Chains, and Their Disease Management and Risks. *Front. Vet. Sci.* 4, 171

177 Clinical and Laboratory Standards Institute (2017) Performance Standards for Antimicrobial Susceptibility Testing*,* 27th Informational Supplement M100

178 Magiorakos, A.P. *et al.* (2012) Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: An international expert proposal for interim standard definitions for acquired resistance. *Clin. Microbiol. Infect.* 18, 268–281

179 Jardine, C.M. *et al.* (2012) Antimicrobial resistance in escherichia coli isolates from raccoons (procyon lotor) in southern ontario, Canada. *Appl. Environ. Microbiol.* 78, 3873–3879

180 Plummer, M. (2003) JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. R Foundation for Statistical Computing

181 Su, Y.S. and Yajima, M. (2012) R2jags: A Package for Running Jags from R. R Foundation for Statistical Computing

182 Goldstein, H. *et al.* (2002) Partitioning Variation in Multilevel Models. *Underst. Stat.* 1, 223–231

183 Pebesma, E.J. (2004) Multivariable geostatistics in S: the gstat package. *Comput. Geosci.* 30, 683–691

184 Anthony, S. *et al.* (2013) A Strategy To Estimate Unknown Viral Diversity in Mammals. *MBio* 4, e00598-13.

185 Chao, A. and Chiu, C. (2016) Species richness: Estimation and comparison. *Wiley StatsRef: Statistics Reference Online* ,1–26

186 Chao, A. *et al.* (2009) Sufficient sampling for asymptotic minimum species richness estimators. *Ecology* 90, 1125–1133

187 IUCN (2017) The IUCN Red List of Threatened Species. International Union for Conservation of Nature and Natural Resources, Cambridge, U.K. http://www.redlist.org/

188 Jetz, W. *et al.* (2004) The scaling of animal space use. *Science* 306, 266–8

189 Ottaviani, D. *et al.* (2006) Body mass as a predictive variable of home-range size among Italian mammals and birds. *J. Zool.* 269, 317–330

190 Lislevand, T. *et al.* (2007) Avian Body Sizes in Relation to Fecundity, Mating System, Display Behaviour, and Resource Sharing. *Ecology* 88, 8888

191 Jones, K.E. *et al.* (2009) PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology* 90, 2648–2648

192 Bahrndorff, S. *et al.* (2016) The Microbiome of Animals: Implications for Conservation Biology. *Int. J. Genomics* DOI:10.1155/2016/5304028

193 Williams, N.J. *et al.* (2011) The prevalence of antimicrobial-resistant Escherichia coli in sympatric wild rodents varies by season and host. *J. Appl. Microbiol.* 110, 962-970

194 Mather, a E. *et al.* (2013) Distinguishable epidemics of multidrug-resistant Salmonella Typhimurium DT104 in different hosts. *Science* 341, 1514–7

195 Spoor, L. *et al.* (2013) Livestock origin for a human pandemic clone of community- associated MRSA. *MBio* 13, e00356-13

196 Ward, M.J. *et al.* (2014) Time-scaled evolutionary analysis of the transmission and antibiotic resistance dynamics of *Staphylococcus aureus* clonal complex 398. *Appl. Environ. Microbiol.* 80, 7275–7282

197 Carroll, D. *et al.* (2015) Antimicrobial Resistance in Wildlife: Implications for Public Health. *Zoonoses Public Health* 62, 534–542

198 Ogutu, J.O. *et al.* (2013) Changing Wildlife Populations in Nairobi National Park and Adjoining Athi-Kaputiei Plains: Collapse of the Migratory Wildebeest. *Open Conservation Biology Journal* 7, 11–26

199 Civitello, D.J. *et al.* (2018) Assessing the direct and indirect effects of food provisioning and nutrient enrichment on wildlife infectious disease dynamics. *Philos. Trans. R. Soc. B.* 373, 20170101

200 Wolters, V. (2000) Invertebrate control of soil organic matter stability. *Biol Fertil Soils* 31, 1–19

201 Potts, S.G. *et al.* (2016) Safeguarding pollinators and their values to human well-being. *Nature* 540, 220–229

202 Frost, L.S. *et al.* (2005) Mobile genetic elements: the agents of open source evolution. *Nat. Rev. Microbiol.* 3, 722–732

203 Lowder, B. V *et al.* (2009) Recent human-to-poultry host jump, adaptation, and pandemic spread of *Staphylococcus aureus*. *Proc. Natl. Acad. Sci. U. S. A.* 106, 19545–19550

204 Gordon, D.M. and Cowling, A. (2003) The distribution and genetic structure of *Escherichia coli* in Australian vertebrates: host and geographic effects. *Microbiology* 149, 3575–3586

205 Chiyo, P.I. *et al.* (2014) The influence of social structure, habitat, and host traits on the transmission of *Escherichia coli* in wild elephants. *PLoS One* 9, e93408

206 VanderWaal, K.L. *et al.* (2014) Quantifying microbe transmission networks for wild and domestic ungulates in Kenya. *Biol. Conserv.* 169, 136–146

207 Bull, C.M. *et al.* (2012) Social networks and the spread of Salmonella in a sleepy lizard population. *Mol. Ecol.* 21, 4386–92

208 Walson, J.L. *et al.* (2001) Carriage of Antibiotic-Resistant Fecal Bacteria in Nepal Reflects Proximity to Kathmandu. *J. Infect. Dis.* 184, 1163–1169

209 Kliemann, B.S. *et al.* (2016) Socioeconomic determinants of antibiotic consumption in the state of são paulo, brazil: The effect of restricting over-the-counter sales. *PLoS One* 11, 1–14

210 Wood, D.E. and Salzberg, S.L. (2014) Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* 15, R46

211 Bankevich, A. *et al.* (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–77

212 Lukjancenko, O. *et al.* (2010) Comparison of 61 sequenced *Escherichia coli* genomes. *Microb. Ecol.* 60, 708–20

213 Larsen, M. V. *et al.* (2014) Benchmarking of methods for genomic taxonomy. *J. Clin. Microbiol.* 52, 1529–1539

214 Larsen, M. V *et al.* (2012) Multilocus sequence typing of total-genome-sequenced bacteria. *J. Clin. Microbiol.* 50, 1355–61

215 Zankari, E. *et al.* (2012) Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* 67, 2640–2644

216 Joensen, K.G. *et al.* (2014) Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J. Clin. Microbiol.* 52, 1501–1510

217 Carattoli, A. *et al.* (2014) In Silico Detection and Typing of Plasmids using PlasmidFinder and Plasmid Multilocus Sequence Typing. *Antimicrob. Agents Chemother.* 58, 3895–3903

218 Mellmann, A. *et al.* (2016) Real-Time Genome Sequencing of Resistant Bacteria Provides Precision Infection Control in an Institutional Setting. *J. Clin. Microbiol.* 54, 2874-2881

219 Ruppitsch, W. *et al.* (2015) Defining and Evaluating a Core Genome Multilocus Sequence Typing Scheme for Whole-Genome Sequence-Based Typing of *Listeria monocytogenes*. *J. Clin. Microbiol.* 53, 2869–76

220 de Been, M. *et al.* (2015) Core Genome Multilocus Sequence Typing Scheme for High- Resolution Typing of *Enterococcus faecium*. *J. Clin. Microbiol.* 53, 3788–97

221 Kluytmans-van den Bergh, M.F.Q. *et al.* (2016) Whole-Genome Multilocus Sequence Typing of Extended-Spectrum-Beta-Lactamase-Producing Enterobacteriaceae. *J. Clin. Microbiol.* 54, 2919–2927

222 Jombart, T. *et al.* (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* 11, 94

223 Jombart, T. (2008) Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24, 1403–1405

224 Jombart, T. (2012) A tutorial for Discriminant Analysis of Principal Components (DAPC) using adegenet. R Foundation for Statistical Computing

225 Kamvar, Z.N. *et al.* (2014) *Poppr* : an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2, e281

226 Weir, B. and Cockerham, C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution (N. Y).* 38, 1358–1370

227 Waples, R.S. (1998) Separating the wheat from the chaff: Patterns of genetic differentiation in high gene flow species. *J. Hered.* 89, 438–450

228 Goudet, J. (2005) Hierfstat, a package for R to compute and test variance components and F-statistics. *Mol. Ecol. Notes* 5, 184–186

229 Warton, D.I. *et al.* (2012) Distance-based multivariate analyses confound location and dispersion effects. *Methods Ecol. Evol.* 3, 89–101

230 Wang, Y. *et al.* (2012) Mvabund- an R package for model-based analysis of multivariate abundance data. *Methods Ecol. Evol.* 3, 471–474

231 Harris, N.C. and Dunn, R.R. (2010) Using host associations to predict spatial patterns in the species richness of the parasites of North American carnivores. *Ecol. Lett.* 13, 1411–1418

232 Kamiya, T. *et al.* (2014) What determines species richness of parasitic organisms? A meta-analysis across animal, plant and fungal hosts. *Biol. Rev.* 89, 123–134

233 Vijaykrishna, D. *et al.* (2008) Evolutionary dynamics and emergence of panzootic H5N1 influenza viruses. *PLoS Pathog.* 4, e1000161

234 Banks, D.J. *et al.* (2002) The fundamental contribution of phages to GAS evolution, genome diversification and strain emergence. *Trends Microbiol.* 10, 515–21

235 Friesen, T.L. *et al.* (2006) Emergence of a new disease as a result of interspecific virulence gene transfer. *Nat. Genet.* 38, 953–956

236 Goldberg, T.L. *et al.* (2007) Patterns of gastrointestinal bacterial exchange between chimpanzees and humans involved in research and tourism in western Uganda. *Biol. Conserv.* 135, 511–517

237 Sheppard, A.E. *et al.* (2016) Nested Russian Doll-Like Genetic Mobility Drives Rapid Dissemination of the Carbapenem Resistance Gene blaKPC. *Antimicrob. Agents Chemother.* 60, 3767–78

238 Rwego, I.B. *et al.* (2008) High rates of *Escherichia coli* transmission between livestock and humans in rural Uganda. *J. Clin. Microbiol.* 46, 3187–91

239 Chaves, L.F. *et al.* (2008) Social Exclusion Modifies Climate and Deforestation Impacts on a Vector-Borne Disease. *PLoS Negl. Trop. Dis.* 2, 1-8

240 Lau, C.L. *et al.* (2010) Climate change, flooding, urbanisation and leptospirosis: Fuelling the fire? *Trans. R. Soc. Trop. Med. Hyg.* 104, 631–638

241 Kowalewski, M.M. *et al.* (2011) Black and gold howler monkeys (*Alouatta caraya*) as sentinels of ecosystem health: Patterns of zoonotic protozoa infection relative to degree of human-primate contact. *Am. J. Primatol.* 73, 75–83

242 Becker, C.G. *et al.* (2015) Deforestation, host community structure, and amphibian disease risk. *Basic Appl. Ecol.* 17, 72–80

243 Gottdenker, N.L. *et al.* (2014) Anthropogenic Land Use Change and Infectious Diseases: A Review of the Evidence. *Ecohealth* 11, 619–632

244 Lawson, B. *et al.* (2015) Citizen Science and Wildlife Disease Surveillance. *Ecohealth* 12, 693–702

245 Gortazar, C. *et al.* (2015) The Wild Side of Disease Control at the Wildlife-Livestock-Human Interface: A Review. *Front. Vet. Sci.* 1, 1–12

246 Becker, D.J. *et al.* (2018) Livestock abundance predicts vampire bat demography, immune profiles, and bacterial infection risk. *Phil. Trans. R. Soc. B.*, 373, 20170089.

247 McCallum, H.I. (2015) Lose biodiversity, gain disease. *Proc. Natl. Acad. Sci.* 112, 8523–8524

248 Suzán, G. *et al.* (2009) Experimental evidence for reduced rodent diversity causing increased hantavirus prevalence. *PLoS One* 4, e5461

249 Keesing, F. *et al.* (2010) Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature* 468, 647–52

250 Young, H.S. *et al.* (2017) Interacting effects of land use and climate on rodent-borne pathogens in central Kenya. *Philos. Trans. R. Soc. B.* 372, 20160116

251 Goddard, M. a *et al.* (2010) Scaling up from gardens: biodiversity conservation in urban environments. *Trends Ecol. Evol.* 25, 90–8

252 Gómez, J.M. *et al.* (2013) Centrality in primate-parasite networks reveals the potential for the transmission of emerging infectious diseases to humans. *Proc. Natl. Acad. Sci. U. S. A.* 110, 7738–41

253 Luis, A.D. *et al.* (2013) A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special? *Proc. R. Soc. B.* 280, 20122753

254 Johnson, P.T.J. *et al.* (2015) Why infectious disease research needs community ecology. *Science* 349, 1259504

Chapter 9

Appendices

type	label
deviceid	
today	
end	
start	
select_one enum	Name of the enumerator
barcode	Site ID
text	
geopoint	Tap to assign GPS
begin group	Animal data
select_one Aclass	Animal classification
select_one yes_no	Will full sampling be performed?
select_one yes_no	Are just faeces available?
end group	
begin group	Background information
decimal	Enter weight
calculate	Max blood collection (ul)
calculate	Pipette multiplication
calculate	MOD of pipette multiplication
calculate	Total volume of blood into PBS
calculate	Total PBS
calculate	Number of times pipette PBS
calculate	Total volume PBS/pipetted blood
calculate	Tube size
note	
note	
note	
select_one yes_no	Is the animal being anaesthetised?
barcode	Anaesthetic barcode
text	
text	Record doses of anaesthetic drugs given
select_one species_name	Select the scientific name
text	Taxonomy to lowest level
select_one idcert	ID certainty
select_one age_class	Age class
select_one sex	Sex
select_one yes_no	Parity
select_one yes_no_unknown	Pregnancy status
select_one yes_no_unknown	Lactating
select_one cond	Condition at sampling
text	Clinical signs
image	Photograph clinical signs
barcode	Sample of lesions
select_one yes_no	Disease suspected
text	Suspected disease
end group	
begin group	Measurements
select_one bcs	Body condition score
select_one yes_no	Is bird a migrant species?
begin group	Measurements
select_one fat	Fat score
decimal	Body length (mm)
decimal	Crown-rump length (mm)
decimal	Ear length (mm)
decimal	Tragus length (mm)
decimal	Head length (mm)
decimal	Beak width (mm)
decimal	Neck circumference (mm)
decimal	Forearm length (mm)
decimal	Wing length (mm)
decimal	Chest circumference (mm)
decimal	Pelvic circumference (mm)
decimal	Tibia length (mm)
decimal	Tarsus length (mm)
decimal	Hind foot length (mm)
decimal	Tail length (mm)
select_one yes_no	Is molt being assessed
text	Molt
text	Secondary molt
text	Tertiary molt
text	Body molt
calculate	Body condition index
image	Full body
image	Full anterior facial
image	Full lateral facial/head
image	Upper dentition
image	Lower dentition
image	Parted pelage on dorsum
image	Parted pelage on vetrum
image	Frontal view of genitalia
image	Ventral view of genitalia
select_one yes_no	Euthanasia
end group	
end group	

Table A.1. Wildlife field sampling form formatted for use with ODK Open Data Kit (ODK) Collect software.

.

type	label
begin group	Samples
select_one yes_no	Is the bird being ringed?
text	Enter ring code
select_one yes_no	Are faeces going for Whole Genome Sequencing?
select_one yes_no	Are faeces going for Campy?
select_one yes_no	Are faeces available?
barcode	Scan first faeces
text	
barcode	Scan second faeces
text	
barcode	Scan campy faeces
text	
barcode	Scan histoplasma faeces
text	
barcode	Scan biobank faeces
text	
barcode	Scan first rectal swab
text	
barcode	Scan second rectal swab
text	
barcode	Scan biobank rectal swab
text	
barcode	Scan first cloacal swab
text	
select_one yes_no	Is a second cloacal swab being taken for WGS?
barcode	Scan second cloacal swab
text	
barcode	Scan biobank cloacal swab
text	
select_one yes_no	Is urine available?
barcode	Scan urine
text	
barcode	Scan oropharyngeal swab
text	
barcode	Scan choanal swab
text	
select_one yes_no	Are ectoparasites present?
barcode	Scan ectoparasites
text	
select_one yes_no	Is there enough blood to collect EDTA?
barcode	Scan EDTA
text	
barcode	Scan serum tube
text	
barcode	Scan first blood smear
text	
barcode	Scan box for first blood smear
text	
integer	Select position in box for first blood smear
barcode	Scan second blood smear
text	
barcode	Scan box for second blood smear
text	
integer	Select position in box for first blood smear
barcode	Scan wing biopsy
text	
barcode	Scan skin biopsy
text	
barcode	Scan hair sample
text	
end group	
begin group	Population track
barcode	Scan population ID
text	
select_one Broost	Type of roost
text	Species present
decimal	Roost size
integer	Roosting count
integer	Exit count
end group	
begin repeat	Bat population samples
barcode	Scan tarp ID
text	
select_one quality	What are the quality of the faeces?
select_one yes_no	Are faeces going for Whole Genome Sequencing?
barcode	Scan urine
text	
barcode	Scan faeces for trizol
text	
barcode	Scan faeces
text	
barcode	Scan second faeces
text	
barcode	Scan first faeces for histoplasma
text	
barcode	Scan second faeces for histoplasma
text	
end repeat	
begin repeat	Avian population samples
barcode	Scan tarp ID
text	
select_one yes_no	Are faeces going for Whole Genome Sequencing?
barcode	Scan faeces
text	
barcode	Scan faeces
text	
barcode	Scan second faeces
text	
end repeat	
text	Enter any comments

type	label
deviceid	
today	
end	
start	
select_one enum	Vet conducting PM
select_one assist	Assitant
barcode	Animal ID
geopoint	Tap to assign GPS
begin group	Animal data
select_one Aclass	Animal classification
end group	
begin group	Initial samples
decimal	Weight
barcode	Scan EDTA
barcode	Scan serum
barcode	Scan first blood smear
barcode	Scan second blood smear
barcode	Scan first oropharyngeal swab
barcode	Scan second oropharyngeal swab
barcode	Scan wing biopsy
barcode	Scan ectoparasites
end group	
begin group	Background information
select_one_external species_name	Select the scientific name
text	Taxonomy to lowest level
select_one idcert	ID certainty
select_one_external age_class	Age class
select_one sex	Sex
select_one yes_no_unknown	Pregnancy status
select_one yes_no_unknown	Lactating
select_one cond	Condition at sampling
text	Clinical signs
select_one yes_no	Disease suspected
text	Suspected disease
begin group	Measurements
select_one_external bcs	Body condition score
decimal	Body length (mm)
decimal	Ear length (mm)
decimal	Tragus length (mm)
decimal	Forearm length (mm)
decimal	Tibia length (mm)
decimal	Hind foot length (mm)
decimal	Tail length (mm)
image	Full body
image	Full anterior facial
image	Full lateral facial/head
image	Parted pelage on dorsum
image	Parted pelage on vetrum
end group	

type	label
begin group	Necropsy
text	Integument lesions
barcode	Scan integument
text	Pectoral muscle lesions
barcode	
	Scan pectoral muscle
text	Ptagium lesions
barcode	Scan ptagium
text	Salivary gland lesions
barcode	Scan salivary glands
barcode	Scan salivary glands
barcode	Scan salivary glands
text	Cavity lesions
text	Diaphgram lesions
barcode	Scan diaphgram
text	Liver lesions
decimal	Liver weight
barcode	Scan liver
barcode	Scan liver
barcode	Scan liver
text	Spleen lesions
decimal	Spleen weight
barcode	Scan spleen
barcode	Scan spleen
barcode	Scan spleen
text	Kidney lesions
decimal	Kidney weight
barcode	Scan kidney
barcode	Scan kidney
barcode	Scan kidney
text	Adrenal lesions
decimal	Adrenal weight
barcode	Scan adrenal
text	Heart lesions
decimal	Heart weight
barcode	Scan heart
text	Lung lesions
decimal	Lung weight
barcode	Scan lung
barcode	Scan lung
barcode	Scan lung
text	Pluck lesions
barcode	Scan pluck
barcode	Scan urine
barcode	Scan urine
barcode	Scan femur
barcode	Scan femur
text	Brain lesions
decimal	Brain weight
barcode	Scan brain
barcode	Scan faeces
barcode	Scan faeces
barcode	Scan faeces
text	Urogenital lesions
barcode	Scan urogenital
barcode	Scan urogenital
text	Stomach lesions
barcode	Scan stomach
text	Ileum lesions
barcode	Scan ileum
barcode	Scan ileum
text	Small intestine lesions
barcode	Scan small intestine
text	Large intesine lesions
barcode	Scan large intestine
text	Scan carcass
end group	
text	Enter any comments

Table A.2. Wildlife necropsy form formatted for use with ODK Open Data Kit (ODK) Collect software.

.

label
S0Q1: Start time
S0Q3: Date
S0Q4: Device ID
S0Q5: Simcard serial number
S0Q6: Device phone number
S0Q7: Recorder Name
S0Q8: Scan household unique barcode on Household Master Sheet
S0Q8.1: The barcode has not scanned correctly. Please go back and try again or enter manually
S1Q1.0: Which types of animals are present in the households today?
S4Q1.1: Which of the following ruminant species are present in this household today?
S4Q1.2: Which of the following monogastric species are present in this household today?
S4Q1.4: Which of the following types of bird are present in this household today?
S4Q1.5: Which of the following types of bird are present in this household today?
S4Q1.6: Which of the following other species are present in this household today?
S4Q1.61: Specify other type(s) of poultry
S4Q1.3: Which of the following other species are present in this household today?
S41.31: Specify other type(s) of animals
S1Q2.0: Use a combination of the farmer's knowledge and your own observations when completing this section
S1Q2.1: Which breed(s) of beef cattle do you own?
S1Q2.2: Number of beef cattle owned
S1Q2.3: How many breeding beef bulls do you own?
S1Q2.4: How many breeding beef cows do you own?
S1Q2.5: How many beef bullocks / steers / heifers do you own?
S1Q2.6: How many beef calves (<6 months old) do you own?
S1Q2.7: Total number of beef animals
S1Q3.1: Which breed(s) of dairy cattle do you own?
S1Q3.2: Number of dairy cattle owned
S1Q3.3: How many breeding dairy bulls do you own?
S1Q3.4: How many lactating dairy cows do you own?
S1Q3.5: How many dry cows do you own?
S1Q3.6: How many heifers do you own?
S1Q3.7: How many dairy calves (<6 months old) do you own?
S1Q3.8: Total number of dairy animals
S1Q4.1: Do you have any dairy breed bullocks / steers that you are rearing for meat?
S1Q4.2: How many dairy bullocks / steers do you own?
S1Q4.3: Total number of dairybeef animals
S1Q4.4: Total number of cattle
S1Q4.5: The total number of cattle present on the premises today is ${s1q4.4_ctl_total}
S1Q4.6: How many separate groups are these animals kept in?
S1Q5.1: Specify breed of meat goat, if known
S1Q5.2: Number of meat goats owned
S1Q5.3: How many breeding bucks do you own?
S1Q5.4: How many breeding does do you own?
S1Q5.5: How many rearing stock (approx 4-12 months old) do you own?
S1Q5.6: How many kids (≤ 3 months old) do you own?
S1:
S1Q5.7: Total number of meat goats
S1Q6.1: Which breed(s) of dairy goats do you own?
S1Q6.2: Number of dairy goats owned
S1Q6.3: How many breeding dairy bucks do you own?
S1Q6.4: How many lactating dairy goats do you own?
S1Q6.5: How many dry/ pregnant goats do you own?
S1Q6.6: How many doelings (not yet kidded) do you own?
S1Q6.7: How many dairy kids(≤ 3 months old) do you own?
S1:
S1Q6.8: Total number of dairy goats
S1Q7.1: Do you have any dairy breed bucks/wethers that you are rearing for meat?
S1Q7.2: How many dairy goats being reared for meat do you own?
S1Q7.3: Total number of dairymeat goats
S1Q7.4: Total number of goats
S1Q7.5: The total number of goats present on the premises today is ${s1q7.4_gt_total}
S1Q7.6: How many separate groups are these animals kept in?
S1Q8.1: Which breed(s) of sheep do you own?
S1Q8.2: Number of sheep owned
S1Q8.3: How many breeding rams do you own?
S1Q8.4: How many breeding ewes do you own?
S1Q8.5: How many fattening lambs (4-12 months old) do you own?
S1Q8.6: How many lambs(≤ 3 months old) do you own?
S1:
S1Q8.7: Total number of sheep
S1Q8.8: The total number of sheep present on the premises today is ${s1q8.7_shp_total}
S1Q8.9: How many separate groups are these animals kept in?
S1Q9.1: Which breed(s) of pig do you own?
S1Q9.2: Number of pigs owned
S1Q9.3: How many breeding boars do you own?
S1Q9.4: How many breeding sows do you own?
S1Q9.5: How many weaners / rearers / finishers do you own (5 weeks old and over) ?
S1Q9.6: How many piglets(<5 weeks old) do you own?
S1:
S1Q9.7: Total number of pigs
S1Q9.8: The total number of pigs present on the premises today is ${s1q9.7_pig_total}
S1Q9.9: How many separate groups are these animals kept in?
S1Q10.1: Which breed(s) of rabbit do you own?
S1Q10.2: Number of rabbits owned
S1Q10.3: How many breeding bucks do you own?
S1Q10.4: How many breeding does do you own?

label
S2: Poultry management
S2Q4.1: Current poultry type
S2Q2.0: Current poultry type
S2Q2.1: Is there purpose-built housing for the ${s2q4.11_curr_pltry}?
S2Q2.2: Poultry cleaning practices
S2Q2.3: How often do you clean the housing?
S2Q2.4: Do you use disinfectant to clean the housing ?
S2Q2.6: Are the ${s2q4.11_curr_pltry} allowed to scavenge?
S2Q2.10: Which of the following do you feed the ${s2q4.11_curr_pltry}?
S2Q2.8: Poultry access to house
S2Q4.11: Do your ${s2q4.11_curr_pltry} share or have access to your general living areas?
S2Q4.3: Do your ${s2q4.11_curr_pltry} have access to food preparation areas?
S3Q3.5: Ruminant value chains
S3: Current ruminant type
S3Q3.6: How many different sources do your ${s3q1.1_curr_rum} come from?
S3Q3.7: Which is the main source of your ${s3q1.1_curr_rum}?
S3Q3.8: Can you tell us the location of the ${s3q1.3_source}?
S3Q3.9: Secondary sources
S3: What is the next source of your ${s3q1.1_curr_rum}?
S3Q4.0: Can you tell us the location of the ${s3q1.6_source}?
S3Q1.8: What do you use your ${s3q1.1_curr_rum} for?
S3Q1.9: What happens to breeding / milking animals at the end of their productive life?
S3Q2.4: Approximately how many ${s3q1.1_curr_rum} have you sold in the past year?
S3Q2.5: Where do you sell your ${s3q1.1_curr_rum}?
S3Q2.6: Which market / abbattoir do you use?
S3: Where do you sell milk from your ${s3q1.1_curr_rum}
S3Q4.3: Monogastric value chains
S3Q4.4: Current monogastric type
S3Q4.5: How many different sources do your ${s3q2.1_curr_mono} come from?
S3: Which is the main source of your ${s3q2.1_curr_mono}?
S3Q4.7: Can you tell us the location of the ${s3q2.3_source}?
S3Q2.8: Secondary sources
S3Q4.8: What is the next source of your ${s3q2.1_curr_mono}?
S3: Can you tell us the location of the ${s3q2.6_source}?
S3Q2.8: What do you use your ${s3q2.1_curr_mono} for?
S3Q2.9: What happens to breeding animals at the end of their productive life?
S3Q2.10: Approximately how many ${s3q2.1_curr_mono} have you sold in the past year?
S3Q2.11: Where do you sell your ${s3q2.1_curr_mono}?
S3Q2.12: Which market / abbattoir do you use?
S3Q2.13: In which subloaction is the community slab that you sell to?
S3Q3.0: Chicken value chains
S3Q3.1: Current chicken type
S3Q3.2: How many different sources do your ${s3q3.1_curr_chkn} chickens come from?
S3Q3.3: Which is the main source of your ${s3q3.1_curr_chkn} chickens?
S3Q3.4: Can you tell us the location of the ${s3q3.3_source}?
S3Q3.5: Which commercial hatchery do your chickens come from?
S3Q3.6: Secondary sources
S3Q3.7: What is the next source of your ${s3q3.1_curr_chkn} chickens?
S3Q3.8: Can you tell us the location of the ${s3q3.7_source}?
S3Q3.9: Which commercial hatchery do your chickens come from?
S3Q3.10: What do you use your ${s3q3.1_curr_chkn} chickens for?
S3Q3.11: What happens to laying birds at the end of their productive life?
S3Q3.12: Approximately how many ${s3q3.1_curr_chkn} chickens have you sold in the past year?
S3Q3.13: How do you sell your ${s3q3.1_curr_chkn} chickens?
S3Q3.14: Where do you sell your ${s3q3.1_curr_chkn} chickens?
S3Q3.15: Where do you sell carcasses from your ${s3q3.1_curr_chkn} chickens?
S3Q3.16: Where do you sell heads and feet from your ${s3q3.1_curr_chkn} chickens?
S3Q3.17: Where do you sell intestines from your ${s3q3.1_curr_chkn} chickens?
S3Q3.18: If selling to a broker:
S3Q3.19: Do you know where the broker/trader sells your ${s3q3.1_curr_chkn} chickens?
S3Q3.20: Specify:
:
S3Q3.15: Where do you sell eggs from your ${s3q3.1_curr_chkn} chickens
:
S3Q4.0: Poultry value chains
S3Q4.1: Current poultry type
S3Q4.11: Current poultry type
S3Q4.2: How many different sources do your ${s3q4.11_curr_pltry} come from?
S3Q4.3: Which is the main source of your ${s3q4.11_curr_pltry}?
S3Q4.4: Can you tell us the location of the ${s3q4.3_source}?
S3Q4.5: Secondary sources
S3Q4.6: What is the next source of your ${s3q4.11_curr_pltry}?
S3Q4.7: Can you tell us the location of the ${s3q4.6_source}?
:
S3Q4.8: What do you use your ${s3q4.11_curr_pltry} for?
S3Q4.9: What happens to laying birds at the end of their productive life?
S3Q4.10: Approximately how many ${s3q4.11_curr_pltry} have you sold in the past year?
S3Q4.11: Where do you sell your ${s3q4.11_curr_pltry} ?
S3Q4.13: Where do you sell eggs from your ${s3q4.11_curr_pltry}
:

label
S4Q3.31: Antibiotic use
S4Q3.32: Which in-water antibiotic do you use?
S4Q3.33: Which in-water anticoccidial do you use?
S4Q3.34: What is the name of the drug?
S4Q3.35: Is the in-water ${s4q3.22_curr_drug} given...?
S4Q3.36: How often do you replace the in-water ${s4q3.22_curr_drug}?
S4Q3.37: Any other relevant information
S4Q3.38: Can you tell us the name of the ${s4q3.22_curr_drug} that you normally use
S4Q4.0: Animals with health problems
S4Q4.1: Current livestock type
S4Q4.2: Have your ${s4q4.1_curr_sp} chickens had any health problems in the last 6 months?
S4Q4.3: Did you use any antibiotics to treat the ${s4q4.1_curr_sp} chickens?
S4Q4.4: What drug formulation(s) did you use?
S4Q4.41: Please specify "other" formulation
S4Q4.5: Antibiotic use
S4Q4.6: Current formulation
S4Q4.71: Which ${s4q4.6_curr_form} antibiotics did you use?
S4Q4.73: Which ${s4q4.6_curr_form} antibiotics did you use?
S4Q4.74: Which ${s4q4.6_curr_form} antibiotics did you use?
S4Q4.75: Which ${s4q4.6_curr_form} antibiotics did you use?
S4Q4.8: What was the name of the drug?
S4Q4.9: Was the ${s4q4.6_curr_form} antibiotic a...?
S4Q4.10: Do you know the dose (mg) or volume (ml) of the ${s4q4.6_curr_form} antibiotic given?
S4Q4.11: What was the amount of ${s4q4.6_curr_form} antibiotic given (per dose)?
S4Q4.12: Select whether dosage reported is in mg or ml.
S4Q4.12B: In what volume of water (litres) was this given?
S4Q4.13: How many ${s4q4.1_curr_sp} were treated?
S4Q4.14: How often was the ${s4q4.6_curr_form} antibiotic given?
S4Q4.15: For how many days was the ${s4q4.6_curr_form} antibiotic used?
S4Q4.16: Were any doses missed?
S4Q4.17: Do you know when the last dose of drug was given ?
S4Q4.18: Any other relevant information
:
S4Q4.19: Do you give your ${s4q4.1_curr_sp} chickens any drugs routinely to prevent disease?
S4Q4.20: Which type(s) of drugs do you routinely use in your ${s4q4.1_curr_sp} chickens?
S4Q4.21: Prophylactic drug use
S4Q4.22: Current drug type
S4Q4.23: What ${s4q4.22_curr_drug} formulation(s) do you use?
S4Q4.24: Antibiotic use
S4Q4.26: Which in-feed antibiotic do you use?
S4Q4.27: Which in-feed anticoccidial do you use?
S4Q4.28: What is the name of the drug?
S4Q4.29: How is the in-feed ${s4q4.22_curr_drug} delivered?
S4Q4.30: Is the in-feed ${s4q4.22_curr_drug} given...?
S4Q4.32: Any other relevant information
S4Q4.33: Antibiotic use
S4Q4.34: Which in-water antibiotic do you use?
S4Q4.35: Which in-water anticoccidial do you use?
S4Q4.36: What is the name of the drug?
S4Q4.37: Is the in-water ${s4q4.22_curr_drug} given...?
S4Q4.38: How often do you replace the in-water ${s4q4.22_curr_drug}?
S4Q4.39: Any other relevant information
S4Q4.40: Can you tell us the name of the ${s4q4.22_curr_drug} that you normally use
S4Q5.0: Animals with health problems
S4Q5.1: Current livestock type
S4Q5.11: Current livestock type
S4Q5.2: Have your ${s4q5.11_curr_sp} had any health problems in the last 6 months?
S4Q5.3: Did you use any antibiotics to treat the ${s4q5.11_curr_sp}?
S4Q5.4: What drug formulation(s) did you use?
S4Q5.41: Please specify "other" formulation
S4Q5.5: Antibiotic use
S4Q5.6: Current formulation
S4Q5.71: Which ${s4q5.6_curr_form} antibiotics did you use?
S4Q5.73: Which ${s4q5.6_curr_form} antibiotics did you use?
S4Q5.74: Which ${s4q5.6_curr_form} antibiotics did you use?
S4Q5.75: Which ${s4q5.6_curr_form} antibiotics did you use?
S4Q5.8: What was the name of the drug?
S4Q5.9: Was the ${s4q5.6_curr_form} antibiotic a...?
S4Q5.10: Do you know the dose (mg) or volume (ml) of the ${s4q5.6_curr_form} antibiotic given?
S4Q5.11: What was the amount of ${s4q5.6_curr_form} antibiotic given (per dose)?
S4Q5.12: Select whether dosage reported is in mg or ml.
S4Q5.12B: In what volume of water (litres) was this given?
S4Q5.13: How many ${s4q5.11_curr_sp} were treated?
S4Q5.14: How often was the ${s4q5.6_curr_form} antibiotic given?
S4Q5.15: For how many days was the ${s4q5.6_curr_form} antibiotic used?
S4Q5.16: Were any doses missed?
S4Q5.17: Do you know when the last dose of drug was given ?
S4Q5.18: Any other relevant information
:
S4Q5.19: Do you give your ${s4q5.11_curr_sp} any drugs routinely to prevent disease?
S4Q5.20: Which type(s) of drugs do you routinely use in your ${s4q5.11_curr_sp}?
S4Q5.21: Prophylactic drug use

S1Q8.9: How many separate groups are these animals kept in?
S1Q9.1: Which breed(s) of pig do you own?
S1Q9.2: Number of pigs owned
S1Q9.3: How many breeding boars do you own?
S1Q9.4: How many breeding sows do you own?
S1Q9.5: How many weaners / rearers / finishers do you own (5 weeks old and over) ?
S1Q9.6: How many piglets(<5 weeks old) do you own?
S1:
S1Q9.7: Total number of pigs
S1Q9.8: The total number of pigs present on the premises today is ${s1q9.7_pig_total}
S1Q9.9: How many separate groups are these animals kept in?
S1Q10.1: Which breed(s) of rabbit do you own?
S1Q10.2: Number of rabbits owned
S1Q10.3: How many breeding bucks do you own?
S1Q10.4: How many breeding does do you own?
S1Q10.5: How many rabbit kittens / young stock do you own?
S1:
S1Q10.6: Total number of rabbits
S1Q10.7: The total number of rabbits present on the premises today is ${s1q10.6_rbt_total}
S1Q10.8: How many separate groups are these animals kept in?
S1Q11.0: Chicken Type
S1Q11.1: Current Chicken Type
S1Q11.2: Number of ${s1q11.1_curr_chkn} chickens owned
S1Q11.3: How many adult cocks do you own?
S1Q11.4: How many laying / breeding hens do you own?
S1Q11.5: How many pullets do you own (6 weeks to point of lay) ?
S1Q11.6: How many broiler chickens do you own?
S1Q11.7: How many chicks or other immature birds do you own?
S1:
S1:
S1Q11.8: How many separate groups are chickens kept in?
S1Q12.0: Other poultry Type
S1Q12.1: Current Type (Other poultry species)
S1Q12.11: Current livestock type
S1Q12.2: Number of ${s1.q12.1.1_curr_sp} owned
S1Q12.3: How many adult males do you own?
S1Q12.4: How many laying / breeding females do you own?
S1Q12.5: How many immature birds do you own?
S1:
S1Q12.6: How many separate groups are ${s1.q12.11_curr_sp} kept in?
S1:
S1Q13: How many horses / donkeys / mules do you own?
S1Q14: How many camels do you own?
S1Q15: How many cats do you own?
S1Q16: How many dogs do you own?
S1Q17: How many ${s41.31_other_spp_own} do you own?
S2Q10.2: Livestock management
S2Q10.3: Current ruminant type
S2Q10.4: Is there purpose-built housing for the ${s2q1.1_curr_rum}?
S2Q10.5: Cleaning practices
S2: How often do you clean the housing?
S2Q10.6: Do you use disinfectant to clean the housing?
S2Q13: Are the ${s2q1.1_curr_rum} allowed to graze / scavenge?
S2Q13: Which of the following do you feed the ${s2q1.1_curr_rum}?
S2Q14: Ruminant access to house
S2Q1.2: Do your ${s2q1.1_curr_rum} share or have access to your general living areas?
S2Q1.4: Do your ${s2q1.1_curr_rum} have access to food preparation areas?
S2Q11.3: Livestock management
S2Q11.4: Current monogastric type
S2Q11.5: Is there purpose-built housing for the ${s2q2.1_curr_mono}?
S2Q11.6: Cleaning practices
S2Q11.7: How often do you clean the housing?
S2: Do you use disinfectant to clean the housing?
S2Q1.7: Are the ${s2q2.1_curr_mono} allowed to graze / scavenge?
S2Q1.7: Which of the following do you feed the ${s2q2.1_curr_mono}?
S2Q1.8: Monogastrics access to house
S2Q2.2: Do your ${s2q2.1_curr_mono} share or have access to your general living areas?
S2Q2.4: Do your ${s2q2.1_curr_mono} have access to food preparation areas?
S2Q1.1: Chicken management
S2Q1.2: Current chicken type
S2Q1.3: Is there purpose-built housing for the ${s2q3.1_curr_chkn} chickens?
S2Q1.4: Chicken cleaning practices
S2Q1.5: How often do you clean the housing?
S2Q1.6: Do you use disinfectant to clean the housing?
S2Q1.7: Are the ${s2q3.1_curr_chkn} chickens allowed to scavenge?
S2: Which of the following do you feed the ${s2q3.1_curr_chkn} chickens?
S2Q2.8: Chickens access to house
S2Q3.1: Do your ${s2q3.1_curr_chkn} chickens share or have access to your general living areas?
S2Q3.3: Do your ${s2q3.1_curr_chkn} have access to food preparation areas?

S3Q4.1: Current poultry type
S3Q4.11: Current poultry type
S3Q4.2: How many different sources do your ${s3q4.11_curr_pltry} come from?
S3Q4.3: Which is the main source of your ${s3q4.11_curr_pltry}?
S3Q4.4: Can you tell us the location of the ${s3q4.3_source}?
S3Q4.5: Secondary sources
S3Q4.6: What is the next source of your ${s3q4.11_curr_pltry}?
S3Q4.7: Can you tell us the location of the ${s3q4.6_source}?
:
S3Q4.8: What do you use your ${s3q4.11_curr_pltry} for?
S3Q4.9: What happens to laying birds at the end of their productive life?
S3Q4.10: Approximately how many ${s3q4.11_curr_pltry} have you sold in the past year?
S3Q4.11: Where do you sell your ${s3q4.11_curr_pltry} ?
S3Q4.13: Where do you sell eggs from your ${s3q4.11_curr_pltry}
:
S3Q16: Are there any other reasons that you keep livestock?
S4Q1: Have any of your animals had any health problems in the last six months?
S4Q2.0: Animals with health problems
S4Q2.1: Current livestock type
S4Q2.2: Have your ${s4q2.1_curr_sp} had any health problems in the last 6 months?
S4Q2.3: Did you use any antibiotics to treat the ${s4q2.1_curr_sp}?
S4Q2.4: What drug formulation(s) did you use?
S4Q2.41: Please specify "other" formulation
S4Q2.5: Antibiotic use
S4Q2.6: Current formulation
S4Q2.7.1: Which ${s4q2.6_curr_form} antibiotics did you use?
S4Q2.7.2: Which ${s4q2.6_curr_form} antibiotics did you use?
S4Q2.7.3: Which ${s4q2.6_curr_form} antibiotics did you use?
S4Q2.7.4: Which ${s4q2.6_curr_form} antibiotics did you use?
S4Q2.7.5: Which ${s4q2.6_curr_form} antibiotics did you use?
S4Q2.8: What was the name of the drug?
S4Q2.9: Was the ${s4q2.6_curr_form} antibiotic a...?
S4Q2.10: Do you know the dose (mg) or volume (ml) of the ${s4q2.6_curr_form} antibiotic given?
S4Q2.11: What was the amount of ${s4q2.6_curr_form} antibiotic given (per dose)?
S4Q2.12: Select whether dosage reported is in mg or ml.
S4Q2.12B: In what volume of water (litres) was this given?
S4Q2.13: How many ${s4q2.1_curr_sp} were treated?
S4Q2.14: How often was the ${s4q2.6_curr_form} antibiotic given?
S4Q2.15: For how many days was the ${s4q2.6_curr_form} antibiotic used?
S4Q2.16: Were any doses missed?
S4Q2.17: Do you know when the last dose of drug was given ?
S4Q2.18: Any other relevant information
S4Q2.19: Do you give your ${s4q2.1_curr_sp} any drugs routinely to prevent disease?
S4Q2.20: Which type(s) of drugs do you routinely use in your ${s4q2.1_curr_sp}?
S4Q2.21: Prophylactic drug use
S4Q2.22: Current drug type
S4Q2.23: Can you tell us the name of the ${s4q2.22_curr_drug} that you normally use
S4:
S4:
S4Q3.0: Animals with health problems
S4Q3.1: Current livestock type
S4Q3.2: Have your ${s4q3.1_curr_sp} had any health problems in the last 6 months?
S4Q3.3: Did you use any antibiotics to treat the ${s4q3.1_curr_sp}?
S4Q3.4: What drug formulation(s) did you use?
S4Q3.41: Please specify "other" formulation
S4Q3.5: Antibiotic use
S4Q3.6: Current formulation
S4Q3.7.1: Which ${s4q3.6_curr_form} antibiotics did you use?
S4Q3.7.3: Which ${s4q3.6_curr_form} antibiotics did you use?
S4Q3.7.4: Which ${s4q3.6_curr_form} antibiotics did you use?
S4Q3.7.5: Which ${s4q3.6_curr_form} antibiotics did you use?
S4Q3.8: What was the name of the drug?
S4Q3.9: Was the ${s4q3.6_curr_form} antibiotic a...?
S4Q3.10: Do you know the dose (mg) or volume (ml) of the antibiotic given?
S4Q3.11: What was the amount of ${s4q3.6_curr_form} antibiotic given (per dose)?
S4Q3.12: Select whether dosage reported is in mg or ml.
S4Q3.12B: In what volume of water (litres) was this given?
S4Q3.13: How many ${s4q3.1_curr_sp} were treated?
S4Q3.14: How often was the ${s4q3.6_curr_form} antibiotic given?
S4Q3.15: For how many days was the ${s4q3.6_curr_form} antibiotic used?
S4Q3.16: Were any doses missed?
S4Q3.17: Do you know when the last dose of drug was given ?
S4Q3.18: Any other relevant information
:
S4Q3.19: Do you give your ${s4q3.1_curr_sp} any drugs routinely to prevent disease?
S4Q3.20: Which type(s) of drugs do you routinely use in your ${s4q3.1_curr_sp}?
S4Q3.21: Prophylactic drug use
S4Q3.22: Current drug type
S4Q3.23: What ${s4q3.22_curr_drug} formulation(s) do you use?
S4Q3.24: Antibiotic use
S4Q3.25: Which in-feed antibiotic do you use?
S4Q3.26: Which in-feed anticoccidial do you use?
S4Q3.27: What is the name of the drug?
S4Q3.28: How is the in-feed ${s4q3.22_curr_drug} delivered?
S4Q3.29: Is the in-feed ${s4q3.22_curr_drug} given...?
S4Q3.30: Any other relevant information

S4Q5.9: Was the ${s4q5.6_curr_form} antibiotic a...?
S4Q5.10: Do you know the dose (mg) or volume (ml) of the ${s4q5.6_curr_form} antibiotic given?
S4Q5.11: What was the amount of ${s4q5.6_curr_form} antibiotic given (per dose)?
S4Q5.12: Select whether dosage reported is in mg or ml.
S4Q5.12B: In what volume of water (litres) was this given?
S4Q5.13: How many ${s4q5.11_curr_sp} were treated?
S4Q5.14: How often was the ${s4q5.6_curr_form} antibiotic given?
S4Q5.15: For how many days was the ${s4q5.6_curr_form} antibiotic used?
S4Q5.16: Were any doses missed?
S4Q5.17: Do you know when the last dose of drug was given ?
S4Q5.18: Any other relevant information
:
S4Q5.19: Do you give your ${s4q5.11_curr_sp} any drugs routinely to prevent disease?
S4Q5.20: Which type(s) of drugs do you routinely use in your ${s4q5.11_curr_sp}?
S4Q5.21: Prophylactic drug use
S4Q5.22: Current drug type
S4Q5.23: What ${s4q5.22_curr_drug} formulation(s) do you use?
S4Q5.24: Antibiotic use
S4Q5.25: Which in-feed antibiotic do you use?
S4Q5.26: Which in-feed anticoccidial do you use?
S4Q5.27: What is the name of the drug?
S4Q5.28: How is the in-feed ${s4q5.22_curr_drug} delivered?
S4Q5.29: Is the in-feed ${s4q5.22_curr_drug} given...?
S4Q5.30: Any other relevant information
S4Q5.31: Antibiotic use
S4Q5.32: Which in-water antibiotic do you use?
S4Q5.33: Which in-water anticoccidial do you use?
S4Q5.34: What is the name of the drug?
S4Q5.35: Is the in-water ${s4q5.22_curr_drug} given...?
S4Q5.36: How often do you replace the in-water ${s4q5.22_curr_drug}?
S4Q5.37: Any other relevant information
S4Q5.38: Can you tell us the name of the ${s4q5.22_curr_drug} that you normally use
S4Q6.0: Animals with health problems
S4Q6.1: Current livestock type
S4Q6.11: Current livestock type
S4Q6.2: Have your ${s4q6.11_curr_sp} had any health problems in the last 6 months?
S4Q6.3: Did you use any antibiotics to treat the ${s4q6.11_curr_sp}?
S4Q6.4: What drug formulation(s) did you use?
S4Q6.41: Please specify "other" formulation
S4Q6.5: Antibiotic use
S4Q6.6: Current formulation
S4Q6.71: Which ${s4q6.6_curr_form} antibiotics did you use?
S4Q6.73: Which ${s4q6.6_curr_form} antibiotics did you use?
S4Q6.74: Which ${s4q6.6_curr_form} antibiotics did you use?
S4Q6.75: Which ${s4q6.6_curr_form} antibiotics did you use?
S4Q6.8: What was the name of the drug?
S4Q6.9: Was the ${s4q6.6_curr_form} antibiotic a...?
S4Q6.10: Do you know the dose (mg) or volume (ml) of the ${s4q6.6_curr_form} antibiotic given?
S4Q6.11: What was the amount of ${s4q6.6_curr_form} antibiotic given (per dose)?
S4Q6.12: Select whether dosage reported is in mg or ml.
S4Q6.12B: In what volume of water (litres) was this given?
S4Q6.13: How many ${s4q6.1_curr_sp} were treated?
S4Q6.14: How often was the ${s4q6.6_curr_form} antibiotic given?
S4Q6.15: For how many days was the ${s4q6.6_curr_form} antibiotic used?
S4Q6.16: Were any doses missed?
S4Q6.17: Do you know when the last dose of drug was given ?
S4Q6.18: Any other relevant information
:
S4Q6.19: Do you give your ${s4q6.11_curr_sp} any drugs routinely to prevent disease?
S4Q6.20: Which type(s) of drugs do you routinely use in your ${s4q6.11_curr_sp}?
S4Q6.21: Prophylactic drug use
S4Q6.22: Current drug type
S4Q6.23: What ${s4q6.22_curr_drug} formulation(s) do you use?
S4Q6.24: Antibiotic use
S4Q6.25: Which in-feed antibiotic do you use?
S4Q6.26: Which in-feed anticoccidial do you use?
S4Q6.27: What is the name of the drug?
S4Q6.28: How is the in-feed ${s4q6.22_curr_drug} delivered?
S4Q6.29: Is the in-feed ${s4q6.22_curr_drug} given...?
S4Q6.30: Any other relevant information
S4Q6.31: Antibiotic use
S4Q6.32: Which in-water antibiotic do you use?
S4Q6.33: Which in-water anticoccidial do you use?
S4Q6.34: What is the name of the drug?
S4Q6.35: Is the in-water ${s4q6.22_curr_drug} given...?
S4Q6.36: How often do you replace the in-water ${s4q6.22_curr_drug}?
S4Q6.37: Any other relevant information
S4Q6.38: Can you tell us the name of the ${s4q6.22_curr_drug} that you normally use
S5Q2.19: You have finished the survey! Thank you very much for your time.
S5Q2.20: End time

Table A.3. Household livestock-keeping questionnaire formatted for use with ODK Open Data Kit (ODK) Collect software.

Ruminants

S0Q1: Start time
S0Q3: Date
S0Q4: Device ID
S0Q5: Simcard serial number
S0Q6: Device phone number
S0Q7: This is the Ruminant sampling questionnaire
S0Q8: Recorder Name
S0Q9: Scan household unique barcode on Household Master Sheet
S1S1: What ruminant animal this?
S2Q1.0: Sampling bovines
S2Q1.3: What is the age of the bovine?
S2Q1.4: Age is given in...
S2Q1.5: What is the breed of the bovine
S2Q1.6: What is the sex of the bovine?
S2Q1.7: What is the body condition score of this bovine?
S2Q1.8: What is the status of the male animal?
S2Q1.9: What is the status of the female animal
S2Q1.10: Has this animal been selected for full genome sequencing?
S2Q2.0: Sampling WITHOUT sequencing
S2Q2.1: Make one faecal swab in Amies transport media and one in Trizol
S2Q2.2: Scan Amies swab 1 barcode
S2Q2.3: Scan Trizol swab 2 barcode
S2Q2.4: Send Amies swab to UoN. Send Trizol swab to ILRI

S2Q3.0: Sampling WITH sequencing
S2Q3.1: Make TWO faecal swabs in Amies transport media and one in Trizol
S2Q3.2: Scan Amies swab 1 barcode
S2Q3.3: Scan Amies swab 2 barcode
S2Q3.4: Scan Trizol swab 3 barcode

:
S2Q4.1: Did you obtain two nasal swabs?
S5Q4.2: Place nasal swab in Trizol
S2Q4.3: Place one nasal swab in Trizol and one in TSB
S2Q4.4: Scan Trizol swab barcode
S2Q4.5: Scan TSB swab barcode
S8Q4.6: Send swab in Trizol to ILRI
S2Q4.7: Send swab in TSB to KEMRI. Send swab in Trizol to ILRI
S2Q4.8: Did you obtain blood samples?
S2Q4.9: Scan EDTA blood sample barcode
S2Q4.10: Scan plain blood sample barcode
S2Q4.11: Send bloods to UoN

S3Q1.0: Sampling caprines
S3Q1.3: What is the age of the caprine?
S3Q1.4: Age is given in...
S3Q1.5: What is the breed of the caprine
S3Q1.6: what is the sex of the caprine

Swine

S0Q1: Start time
S0Q3: Date
S0Q4: Device ID
S0Q5: Simcard serial number
S0Q6: Device phone number
S0Q7: This is the Swine sampling questionnaire
S0Q8: Recorder Name
S0Q9: Scan household unique barcode on Household
S1Q1.1: Which of these best describes the animals in this group?
S1Q2.1: How many pigs in this group?
S1Q2.2: How many pigs are we sampling from this group?
S1Q2.3: How many pigs are we sending for full genome
S2Q1.0: Sampling pigs
S2Q1.3: How old is this animal?
S2Q1.4: Age is given in...
S2Q1.5: What is the breed of the swine
S2Q1.6: what is the sex of the swine
S2Q1.7: What is the body condition score of this swine
S2Q1.8: what is the status of the male animal
S2Q1.9: what is the status of the female animal
S2Q1.10: Has this animal been selected for full genome sequencing?
S2Q2.0: Sampling WITHOUT sequencing
S2Q2.1: Make one faecal swab in Amies transport media and one in Trizol
S2Q2.2: Scan Amies swab 1 barcode
S2Q2.3: Scan Trizol swab 2 barcode
S2Q2.4: Send Amies swab to UoN. Send Trizol swab to ILRI

S2Q3.0: Sampling WITH sequencing
S2Q3.1: Make TWO faecal swabs in Amies transport media and one in Trizol
S2Q3.2: Scan Amies swab 1 barcode
S2Q3.3: Scan Amies swab 2 barcode
S2Q3.4: Scan Trizol swab 3 barcode

:
S4Q4.1: Did you obtain two nasal swabs?
S5Q4.2: Place nasal swab in Trizol
S5Q4.3: Place one nasal swab in Trizol and one in TSB
S6Q4.4: Scan Trizol swab barcode
S7Q4.5: Scan TSB swab barcode
S8Q4.6: Send swab in Trizol to ILRI
S8Q4.7: Send swab in TSB to KEMRI. Send swab in Trizol to ILRI
S9Q4.8: Did you obtain blood samples?
S10Q4.9: Scan EDTA blood sample barcode
S11Q4.10: Scan plain blood sample barcode
S12Q4.11: Send bloods to UoN

S5Q1: End time

Rabbits

S0Q1: Start time
S0Q3: Date
S0Q4: Device ID
S0Q5: Simcard serial number
S0Q6: Device phone number
S0Q7: This is the Rabbit sampling questionnaire
S0Q8: Recorder Name
S0Q9: Scan household unique barcode on Household
S1Q1.1: Which of these best describes the animals in this group?
S1Q2.1: How many rabbits in this group?
S1Q2.2: How many rabbits are we sampling from this
S1Q2.3: How many rabbits are we sending for full
S2Q1.0: Sampling rabbits
S2Q1.3: How old is this animal?
S2Q1.4: Age is given in...
S2Q1.5: What is the breed of the rabbit
S2Q1.6: what is the sex of the rabbit
S2Q1.7: What is the body condition score of this rabbit
S2Q1.8: what is the status of the male animal
S2Q1.9: what is the status of the female animal
S2Q1.10: Has this animal been selected for full genome sequencing?
S2Q2.0: Sampling WITHOUT sequencing
S2Q2.1: Make one faecal swab in Amies transport media and one in Trizol
S2Q2.2: Scan Amies swab 1 barcode
S2Q2.3: Scan Trizol swab 2 barcode
S2Q2.4: Send Amies swab to UoN. Send Trizol swab to ILRI

S2Q3.0: Sampling WITH sequencing
S2Q3.1: Make TWO faecal swabs in Amies transport media and one in Trizol
S2Q3.2: Scan Amies swab 1 barcode
S2Q3.3: Scan Amies swab 2 barcode
S2Q3.4: Scan Trizol swab 3 barcode

:
S4Q4.1: Did you obtain two nasal swabs?
S5Q4.2: Place nasal swab in Trizol
S5Q4.3: Place one nasal swab in Trizol and one in TSB
S6Q4.4: Scan Trizol swab barcode
S7Q4.5: Scan TSB swab barcode
S8Q4.6: Send swab in Trizol to ILRI
S8Q4.7: Send swab in TSB to KEMRI. Send swab in Trizol to ILRI

S5Q1: End time

Ruminants Contd.

S3Q1.7: What is the body condition score of this caprine
S3Q1.8: What is the status of the male animal?
S3Q1.9: What is the status of the female animal
S3Q1.10: Has this animal been selected for full genome sequencing?
S3Q2.0: Sampling WITHOUT sequencing
S3Q2.1: Make one faecal swab in Amies transport media and one in Trizol
S3Q2.2: Scan Amies swab 1 barcode
S3Q2.3: Scan Trizol swab 2 barcode
S3Q2.4: Send Amies swab to UoN. Send Trizol swab to ILRI

S3Q3.0: Sampling WITH sequencing
S3Q3.1: Make TWO faecal swabs in Amies transport media and one in Trizol
S3Q3.2: Scan Amies swab 1 barcode
S3Q3.3: Scan Amies swab 2 barcode
S3Q3.4: Scan Trizol swab 3 barcode

:

S3Q4.1: Did you obtain two nasal swabs?
S5Q4.2: Place nasal swab in Trizol
S3Q4.3: Place one nasal swab in Trizol and one in TSB
S3Q4.4: Scan Trizol swab barcode
S3Q4.5: Scan TSB swab barcode
S8Q4.6: Send swab in Trizol to ILRI
S3Q4.7: Send swab in TSB to KEMRI. Send swab in Trizol to ILRI
S3Q4.8: Did you obtain blood samples?
S3Q4.9: Scan EDTA blood sample barcode
S3Q4.10: Scan plain blood sample barcode
S3Q4.11: Send bloods to UoN

S4Q1.0: Sampling ovines
S4Q1.3: What is the age of the ovine?
S4Q1.4: Age is given in...
S4Q1.5: What is the breed of the ovine
S4Q1.6: what is the sex of the ovine
S4Q1.7: What is the body condition score of this ovine
S4Q1.8: What is the status of the male animal?
S4Q1.9: What is the status of the female animal
S4Q1.10: Has this animal been selected for full genome sequencing?
S4Q2.0: Sampling WITHOUT sequencing
S4Q2.1: Make one faecal swab in Amies transport media and one in Trizol
S4Q2.2: Scan Amies swab 1 barcode
S4Q2.3: Scan Trizol swab 2 barcode
S4Q2.4: Send Amies swab to UoN. Send Trizol swab to ILRI

S4Q3.0: Sampling WITH sequencing
S4Q3.1: Make TWO faecal swabs in Amies transport media and one in Trizol
S4Q3.2: Scan Amies swab 1 barcode
S4Q3.3: Scan Amies swab 2 barcode
S4Q3.4: Scan Trizol swab 3 barcode

:

S4Q4.1: Did you obtain two nasal swabs?
S5Q4.2: Place nasal swab in Trizol
S5Q4.3: Place one nasal swab in Trizol and one in TSB
S6Q4.4: Scan Trizol swab barcode
S7Q4.5: Scan TSB swab barcode
S8Q4.6: Send swab in Trizol to ILRI
S8Q4.7: Send swab in TSB to KEMRI. Send swab in Trizol to ILRI
S9Q4.8: Did you obtain blood samples?
S10Q4.9: Scan EDTA blood sample barcode
S11Q4.10: Scan plain blood sample barcode
S12Q4.11: Send bloods to UoN

Poultry

S0Q1: Start time
S0Q3: Date
S0Q4: Device ID
S0Q5: Simcard serial number
S0Q6: Device phone number
S0Q7: This is the Poultry sampling questionnaire
S0Q8: Recorder Name
S0Q9: Scan household unique barcode on Household
S1Q1.1: What type of poultry is this?
S1Q1.2: What type of chickens are these?
S1Q1.3: What species of poultry is this?
S1Q2.0: Are the birds in this group...?
S1Q2.1: (Approximately) how many birds in this group?
S1Q2.2: How many birds are we sampling from this
S1Q2.3: How many birds are we sending for full genome
S2Q1.0: Commercial Layers
S2Q1.1: Are these birds...?
S2Q1.2: How many weeks in lay are these birds?
S2Q1.3: How many weeks of age are these birds?
S2Q1.4: Where did the birds come from?
S2Q2.0: Sampling layers WITHOUT sequencing
S2Q2.1: Make one faecal swab in Amies transport media and one in Trizol
S2Q2.2: Scan Amies swab 1 barcode
S2Q2.3: Scan Trizol swab 2 barcode
S2Q2.4: Send Amies swab to UoN. Send Trizol swab to ILRI
S2Q2.5: Make one oropharyngeal swab in Trizol and one in TSB
S2Q2.6: Scan TSB oro-pharyngeal swab barcode
S2Q2.7: Scan Trizol oro-pharyngeal swab barcode
S2Q2.8: Send OP swab in TSB to KEMRI. Send swab in Trizol to ILRI

S2Q3.0: Sampling layers WITH sequencing
S2Q3.1: Make TWO faecal swabs in Amies transport media and one in Trizol
S2Q3.2: Scan Amies swab 1 barcode
S2Q3.3: Scan Amies swab 2 barcode
S2Q3.4: Scan Trizol swab 3 barcode

S2Q3.6: Make one oropharyngeal swab in Trizol and one in TSB
S2Q3.7: Scan TSB oro-pharyngeal swab barcode
S2Q3.8: Scan Trizol oro-pharyngeal swab barcode
S2Q3.9: Send OP swab in TSB to KEMRI. Send swab in Trizol to ILRI

S3Q1.0: Commercial Broilers
S3Q1.1: How many weeks of age are these birds?
S3Q1.2: Where did the birds come from?
S3Q2.0: Sampling broilers WITHOUT sequencing
S3Q2.1: Make one faecal swab in Amies transport media
S3Q2.2: Scan Amies swab 1 barcode
S3Q2.3: Scan Trizol swab 2 barcode
S3Q2.4: Send Amies swab to UoN. Send Trizol swab to
S3Q2.5: Make one oropharyngeal swab in Trizol and one
S3Q2.6: Scan TSB oro-pharyngeal swab barcode
S3Q2.7: Scan Trizol oro-pharyngeal swab barcode
S3Q2.8: Send OP swab in TSB to KEMRI. Send swab in Trizol to ILRI

S3Q3.0: Sampling broilers WITH sequencing
S3Q3.1: Make TWO faecal swabs in Amies transport media and one in Trizol
S3Q3.2: Scan Amies swab 1 barcode
S3Q3.3: Scan Amies swab 2 barcode
S3Q3.4: Scan Trizol swab 3 barcode

Poultry Contd.

S3Q3.6: Make one oropharyngeal swab in Trizol and one in TSB
S3Q3.7: Scan TSB oro-pharyngeal swab barcode
S3Q3.8: Scan Trizol oro-pharyngeal swab barcode
S3Q3.9: Send OP swab in TSB to KEMRI. Send swab in Trizol to ILRI

S4Q1.0: Dual-purpose chickens or other poultry species (uniform age)
S4Q1.1: Are these birds...?
S4Q1.2: How many weeks in lay are these birds?
S4Q1.1: How many weeks of age are these birds?
S4Q1.2: Where did the birds come from?
S4Q2.0: Sampling birds WITHOUT sequencing
S4Q2.1: Make one faecal swab in Amies transport media and one in Trizol
S4Q2.2: Scan Amies swab 1 barcode
S4Q2.3: Scan Trizol swab 2 barcode
S4Q2.4: Send Amies swab to UoN. Send Trizol swab to ILRI
S4Q2.5: Make one oropharyngeal swab in Trizol and one
S4Q2.6: Scan TSB oro-pharyngeal swab barcode
S4Q2.7: Scan Trizol oro-pharyngeal swab barcode
S4Q2.8: Send OP swab in TSB to KEMRI. Send swab in

S4Q3.0: Sampling birds WITH sequencing

S4Q3.2: Scan Amies swab 1 barcode
S4Q3.3: Scan Amies swab 2 barcode
S4Q3.4: Scan Trizol swab 3 barcode

S4Q3.6: Make one oropharyngeal swab in Trizol and one in TSB
S4Q3.7: Scan TSB oro-pharyngeal swab barcode
S4Q3.8: Scan Trizol oro-pharyngeal swab barcode
S4Q3.9: Send OP swab in TSB to KEMRI. Send swab in Trizol to ILRI

S5Q1.0: Indigenous chickens or other poultry species (mixed age groups)
S5Q2.0: Sampling birds WITHOUT sequencing
S5Q2.1: What is the sex of this bird?
S5Q2.2: What stage of production is this bird?
S5Q2.3: How many weeks of age is this bird?
S4Q1.2: Where did the birds come from?
S5Q2.4: Make one faecal swab in Amies transport media and one in Trizol
S5Q2.5: Scan Amies swab 1 barcode
S5Q2.6: Scan Trizol swab 2 barcode
S5Q2.7: Send Amies swab to UoN. Send Trizol swab to ILRI
S5Q2.8: Make one oropharyngeal swab in Trizol and one in TSB
S5Q2.9: Scan TSB oro-pharyngeal swab barcode
S5Q2.10: Scan Trizol oro-pharyngeal swab barcode
S5Q2.11: Send OP swab in TSB to KEMRI. Send swab in Trizol to ILRI

S5Q3.0: Sampling birds WITH sequencing
S5Q3.1: What is the sex of this bird?
S5Q3.2: What stage of production is this bird?
S5Q3.3: How many weeks of age is this bird?
S4Q3.4: Where did the birds come from?
S5Q3.5: Make TWO faecal swabs in Amies transport
S5Q3.6: Scan Amies swab 1 barcode
S5Q3.7: Scan Amies swab 2 barcode
S5Q3.8: Scan Trizol swab 3 barcode

S5Q3.10: Make one oropharyngeal swab in Trizol and one
S5Q3.11: Scan TSB oro-pharyngeal swab barcode
S5Q3.12: Scan Trizol oro-pharyngeal swab barcode
S5Q3.13: Send OP swab in TSB to KEMRI. Send swab in

Table A.4. Livestock sampling form for ruminants, swine, rabbits and poultry, formatted for use with ODK Open Data Kit (ODK) Collect software.

label
S0Q1: Start time
S0Q2: End time
S0Q3: Date
S0Q4: Device ID
S0Q5: Simcard serial number
S0Q6: Device phone number
S0Q7: Recorder Name
S0Q8: This questionnaire is being conducted in...?
S0Q9: Scan household unique barcode on Household Master Sheet
S0Q91: The barcode has not scanned correctly. Please go back and try again or enter manually
S1Q11: Select the first letter of the current Sublocation
S1Q12: Select the current sublocation
S1Q2: Is this a livestock-owning household?
S1Q3: Select a household member to start the livestock questionnaire
S2Q1: What is the first name of the household head?
S2Q2: Which gender is ${s2q1_hh_head_id}?
S2Q3: What is the total number of family members in this household?
S2Q41: How many adult (18 and over) family members live in this household?
S2Q42: How many of the adult family members of this household are present today?
S2Q51: Is the household head present today?
S2Q52: Did ${s2q1_hh_head_id} provide a faecal sample?
S2Q53: Did ${s2q1_hh_head_id} complete a consent form?
S2Q53B: Unfortunately we cannot accept a faecal sample without a completed consent form
S2Q54: How old is ${s2q1_hh_head_id}?
S2Q55: What education has ${s2q1_hh_head_id} had?
S2Q56: What ethnicity is ${s2q1_hh_head_id}?
S2Q57: Which tribe is ${s2q1_hh_head_id}?
S2Q58: What is ${s2q1_hh_head_id}'s occupation?
S2Q59: Does ${s2q1_hh_head_id} work with any of the following outside the household?
S2Q6: Number of adult family members absent today
S2Q60: Adult family members absent today
S2Q61: What is the first name of the absent family member
S2Q62: Which gender is ${s2q61_adlt_abs_id}?
S2Q63: What relation is ${s2q61_adlt_abs_id} to ${s2q1_hh_head_id}?
S2Q64: What relation is ${s2q61_adlt_abs_id} to ${s2q1_hh_head_id}?
S2Q65: Did ${s2q61_adlt_abs_id} provide a faecal sample?
S2Q66: Did ${s2q61_adlt_abs_id} complete a consent form?
S2Q66B: Unfortunately we cannot accept a faecal sample without a completed consent form
S2Q67: How old is ${s2q61_adlt_abs_id}?
S2Q68: What education has ${s2q61_adlt_abs_id} had?
S2Q69: What ethnicity is ${s2q61_adlt_abs_id}?
S2Q610: Which tribe is ${s2q61_adlt_abs_id}?
S2Q611: What is ${s2q61_adlt_abs_id}'s occupation?
S2Q612: Does ${s2q61_adlt_abs_id} work with any of the following outside the household?
S2Q7: How many children (under 18) live in this household?
S2Q612: How many of the children in this household are present today?
S2Q8: Number of children present today
S2Q81: Number of children absent today
S2Q82: Children absent today
S2Q90: What is the first name of the absent child
S2Q91: Which gender is ${s2q91_chl_abs_id}?
S2Q92: What relation is ${s2q91_chl_abs_id} to ${s2q1_hh_head_id}?
S2Q93: What relation is ${s2q91_chl_abs_id} to ${s2q1_hh_head_id}?
S2Q99: Did ${s2q91_chl_abs_id} provide a faecal sample?
S2Q96: Did ${s2q91_chl_abs_id} complete an assent form?
S2Q96B: Unfortunately we cannot accept a faecal sample without a completed consent form
S2Q94: How old is ${s2q91_chl_abs_id}?
S2Q97: What education has ${s2q91_chl_abs_id} had?
S2Q98: What ethnicity is ${s2q91_chl_abs_id}?
S2Q910: Which tribe is ${s2q91_chl_abs_id}?
S2Q911: Does ${s2q91_chl_abs_id} attend school?
S2Q912: Does ${s2q91_chl_abs_id} eat food provided at school?
S2Q913: Does ${s2q91_chl_abs_id} eat any of these from the school?

label
S2: Do you employ any staff?
S2Q10: Staff employed
S2Q111: How many indoor staff (maids / cleaning staff / housekeepers / cook/nanny, etc) do you employ?
S2Q112: Number of indoor staff
S2Q113: How many guards do you employ?
S2Q114: Number of guards
S2Q115: How many farmhands / gardeners do you employ?
S2Q116: Number of farmhands
S2Q12: Number of staff employed
S2Q121: How many staff are present today?
S2Q122: Number of staff present
S2Q13: Do any other non-family members (eg. friends, tenants) live in the household?
S2Q131: How many non-family members live in the house?
S2Q132: How many of them are present today?
S2Q133: Number of other adults present today
S2Q14: Total number of people present in the household today
S2Q15: The total number of people present in the household today is ${s2q14_sum_pres}. Is this correct?
S3Q0: The next section of the questionnaire relates to your house and land
S3Q1: Which of these best describes the housing type?
S3Q2: Do you own the house / apartment?
S3Q3: Do you own the land on which the house is built?
S3Q4: Was the house...?
S3Q5: How many hectares of land do you have at this location?
S3Q4: Which of these best describes your situation?
S3Q5: How much rent do you pay per month?
S3Q6: Do you own and/or rent any land elsewhere?
S3Q7: Land owned or rented
S3Q8: Current land status
S3Q9: Where is the land that you ${s3q91_land_curr}?
S3Q91: Location of land
S3Q10: Current land location
S3Q11: Location of land
S3Q91: What is the name of sublocation / region where your land is located?
S3Q7: How many hectares of land do you ${s3q91_land_curr} ${s3q111_land_loc_curr} Nairobi?
S3Q8: How many hectares of the land that you ${s3q91_land_curr} ${s3q111_land_loc_curr} Nairobi is used for agriculture?
S3Q9: Do you keep any livestock on the land that you ${s3q91_land_curr} ${s3q111_land_loc_curr} Nairobi?
S3Q10: Which of these livestock do you keep on the land that you ${s3q91_land_curr} ${s3q111_land_loc_curr} Nairobi?
S3Q8: How many rooms in your house?
S3Q114: How many bedrooms in your house?
S3Q10: How many rooms do you rent out?
S3Q11: What is the floor made from?
S3Q12: What are the walls made from?
S3Q13: What is the roof made from?
S3Q15: Transport
S3Q15: Does anyone in the household own a car or truck?
S3Q16: Does anyone in the household own a motorbike or scooter?
S3Q171: Does anyone in the household own a bicycle?
S3Q172: Does anyone in the household own an animal-drawn cart?
S3Q17: Electrical items
S3Q171: Does the household have electricity?
S3Q172: Does the household have a solar panel?
S3Q172: Do you own a refrigerator?
S3Q173: Do you own a television?
S3Q183: Do you own a radio?
S3Q183: Personal belongings
S3Q184: Does anyone in the household own a watch?
S3Q185: Does anyone in the household own a mobile phone?

label
S3Q191: Where does your household do most of its cooking?
S3Q192: Which sources of cooking fuel do you use?
S3Q25: What is your main source of lighting?
S3Q26: What is your main source of drinking water?
S3Q27: What is your main source of non-drinking water?
S3Q27: Do you have a water tank?
S3Q28: Do you use any water treatments for drinking water?
S3Q29: How would you rate the water quality?
S3Q30: Is the water supply always regular?
S3Q31: What type of toilet facility does your family use?
S3Q32: Where is the facility located?
S3Q33: Do you share this facility with other families?
S3Q34: What happens to waste water?
S3Q35: How do you dispose of household waste?
S3Q36: What do you do with animal waste?
S4Q11: Do you see any of the following types of wild animal around your property?
S4Q12: Which of the following carnivores have you seen around your property?
S4Q13: Which of the following primates have you seen around your property?
S4Q20: Wildlife encounters
S4Q21: Current wildlife type
S4Q22: Where do you see ${s4q21_curr_wl}s?
S4Q23: Do you ever see ${s4q21_curr_wl}s inside the house?
S4Q24: Do you ever see ${s4q21_curr_wl}s inside the kitchen?
S4Q25: Approximately how many ${s4q21_curr_wl}s do you see inside the house and/or kitchen?
S4Q26: per...
S4Q27: Do you ever see ${s4q21_curr_wl}s inside the animal housing?
S4Q28: Approximately how many ${s4q21_curr_wl}s do you see in the animal housing?
S4Q29: per...
S4Q210: Approximately how many ${s4q21_curr_wl}s do you see outside?
S4Q211: per...
S4Q30: How often do you see ${s4q21_curr_wl}s?
S4Q31: Do you think ${s4q21_curr_wl}s cause your family any of the following problems?
S4Q32: Do you think ${s4q21_curr_wl}s cause your family any of the following problems?
S4Q33: Are there any measures you use to try and control ${s4q21_curr_wl}s?
S4Q34: What type of poison do you use?
S4Q35: How many ${s4q21_curr_wl}s do you trap?
S4Q36: Number of ${s4q21_curr_wl}s trapped...
S4Q37: per...
S4Q9: Has anyone in the household ever been bitten by a ${s4q21_curr_wl}?
S5Q1: You have finished the questionnaire! Thank you very much.
S5Q2: If any absent family members have provided consent and faecal samples, collect them now
S5Q20: Household head absent - faecal sample left
S5Q21: You now need to make one swab from ${s2q1_hh_head_id}'s stool sample and place in Amies transport media. Make a second swab and place in Trizol
S5Q22: Label and scan the faecal sample pot for ${s2q1_hh_head_id}
S5Q22B: The barcode has not scanned correctly. Please go back and try again or enter manually
S5Q23: Label and scan the Amies faecal swab ${s2q1_hh_head_id}
S5Q23B: The barcode has not scanned correctly. Please go back and try again or enter manually
S5Q24: Label and scan the Trizol faecal swab for ${s2q1_hh_head_id}
S5Q24B: The barcode has not scanned correctly. Please go back and try again or enter manually
S5Q30: Adult family members absent today
S5Q31: Current adult family member absent today
S5Q32: Current adult family member sample provided?
S5Q33: Current adult family member consent provided?
S5Q34: You now need to make one swab from ${s5q31_adlt_abst_no}'s stool sample and place in Amies transport media. Make a second swab and place in Trizol
S5Q35: Label and scan the faecal sample pot for ${s5q31_adlt_abst_no}
S5Q35B: The barcode has not scanned correctly. Please go back and try again or enter manually
S5Q36: Label and scan the faecal swab for ${s5q31_adlt_abst_no}
S5Q36B: The barcode has not scanned correctly. Please go back and try again or enter manually
S5Q37: Label and scan the Trizol faecal swab for ${s5q31_adlt_abst_no}
S5Q37B: The barcode has not scanned correctly. Please go back and try again or enter manually
S5Q40: Child family members absent today
S5Q41: Current child family member absent today
S5Q42: Current child family member sample provided?
S5Q43: Current child family member consent provided?
S5Q44: You now need to make one swab from ${s5q41_child_abst_no}'s stool sample and place in Amies transport media. Make a second swab and place in Trizol
S5Q45: Label and scan the faecal sample pot for ${s5q41_child_abst_no}
S5Q45B: The barcode has not scanned correctly. Please go back and try again or enter manually
S5Q46: Label and scan the faecal swab for ${s5q41_child_abst_no}
S5Q46B: The barcode has not scanned correctly. Please go back and try again or enter manually
S5Q47: Label and scan the Trizol faecal swab for ${s5q41_child_abst_no}
S5Q47B: The barcode has not scanned correctly. Please go back and try again or enter manually
S5Q51: Capture the GPS of the current location.
S5Q52: Confirm the GPS co-ordinates

Table A.5. Household questionnaire, formatted for use with ODK Open Data Kit (ODK) Collect software.

Chapter 10

Related Publications

1. Hassell, J. M., Begon, M., Ward, M. J., & Fèvre, E. M. (2017). Urbanization and disease emergence: Dynamics at the wildlife–livestock–human interface. Trends in Ecology & Evolution, 32, 55–67.

CellPress

Review

Urbanization and Disease Emergence: Dynamics at the Wildlife–Livestock–Human Interface

James M. Hassell,[1,2] Michael Begon,[3] Melissa J. Ward,[4] and Eric M. Fèvre[1,2,*]

Urbanization is characterized by rapid intensification of agriculture, socioeconomic change, and ecological fragmentation, which can have profound impacts on the epidemiology of infectious disease. Here, we review current scientific evidence for the drivers and epidemiology of emerging wildlife-borne zoonoses in urban landscapes, where anthropogenic pressures can create diverse wildlife–livestock–human interfaces. We argue that these interfaces represent a critical point for cross-species transmission and emergence of pathogens into new host populations, and thus understanding their form and function is necessary to identify suitable interventions to mitigate the risk of disease emergence. To achieve this, interfaces must be studied as complex, multihost communities whose structure and form are dictated by both ecological and anthropological factors.

Trends

Urbanization can create diverse wildlife–livestock–human interfaces.

Interfaces represent a critical point for cross-species transmission and emergence of pathogens.

Interfaces should be studied as complex, multihost communities.

Molecular epidemiology can add real-world complexity to the study of disease emergence.

Emerging Diseases in Changing Landscapes

Emerging infectious diseases (**EIDs**) (see Glossary) are recognized as pathogens 'whose incidence in host populations has increased within the past two decades or threatens to increase in the near future' [1]. As well as describing the spread of newly evolved or previously undetected pathogens, pathogens that are increasing their geographic spread, increasing their impact, changing their clinical presentation or moving into human hosts for the first time, the term emergence can also be used to describe the reappearance (or re-emergence) of a known infection after a decline in incidence [1]. It is estimated that between 60 and 80% of newly emerging infections are zoonotic in origin and thus are (at least initially) dependent on an animal **reservoir** for survival [2,3]. Of these emerging zoonoses, at least 70% have a wildlife origin, with cross-species spread and onward transmission representing a natural response to the evolutionary pressures of pathogen ecology [3,4]. Although both wildlife and domesticated animal reservoirs can be considered important sources of EIDs, it is the anthropogenic influence on ecological systems that dictates the level of risk that operates at the **interface** between humans and animals in zoonotic disease emergence.

The impact of humans on the ecosystems within which they exist have occurred for as long as there have been humans. However, over the past 10 000 years, human–ecosystem interactions have become increasingly profound following a series of chronological transitions: (i) the establishment of local settlements, agriculture, and domestication of livestock; (ii) regional contact through trade; (iii) intercontinental exploration, imperialism, and industrialization; and

[1]Institute of Infection and Global Health, The University of Liverpool, Leahurst Campus, Chester High Road, Neston, CH64 7TE, UK
[2]International Livestock Research Institute, Nairobi, Kenya
[3]Institute of Integrative Biology, The University of Liverpool, Liverpool L69 7ZB, UK
[4]Centre for Immunity, Infection and Evolution, University of Edinburgh, Edinburgh, UK

*Correspondence: eric.fevre@liverpool.ac.uk (E.M. Fèvre).

(iv) globalization, urbanization, and climate change [5]. Current levels of human–ecosystem interaction, driven by increased environmental encroachment and **land-use change** (exploitation of natural resources and agricultural practices), and environmental effects such as climate change, will result in habitat alteration and changes in species assemblage and contact rates that promote the emergence of zoonotic disease. Spread and persistence of newly emerged (or re-emerged) pathogens can then be perpetuated by a combination of factors including expanding global human populations and urbanization, international trade and travel, intensive livestock keeping systems, proliferation of reservoir populations, and antimicrobial drug use [4,6–8]. Land-use change, through population-driven anthropogenic influences such as forestry, mining, agriculture, and urban and industrial development, is frequently associated with disease emergence [9,10].

Urbanization can be considered a key driver of land-use change that is likely to increase at an unprecedented rate in the coming decades, particularly in developing countries, where as much of 90% of population growth is projected to occur in cities [11,12]. Human population density and growth are significant predictors of historical EID events, and thus urbanization is likely to have a profound effect on public health as rural pathogens adapt to urban conditions, and other pathogens emerge (or re-emerge) in urban areas [3]. Human factors such as population density, migration, trade, sanitation, and access to clean water can promote the transmission of pathogens and alter vector dynamics, while social factors that drive health inequality (socioeconomic status, housing, race, ethnicity, gender, and education) also influence the epidemiology of infectious disease in urban areas [13,14] (Figure 1). For cities in developing countries, the epidemiological effects of these factors are often concentrated in informal settlements, where population growth and density is highest [14]. In this review, we focus on rapid urbanization (predominantly a feature of developing countries) as a driver of disease emergence, and use it to explore how anthropogenic changes are driving interactions and the potential for disease emergence between sympatric wildlife, livestock, and humans.

Urbanization and Disease Emergence

Spatial overlap between hosts, and overlap in vector ranges are key requirements for the emergence of directly transmitted and vector-borne pathogens, respectively. As such, in order to investigate the conditions in which urbanization might lead to the emergence of zoonotic disease across species, and thus risk factors for transmission to humans, it is necessary to simplify the complexity of urban systems by considering them as a network of interfaces across which pathogens can be transmitted; the physical interfaces at which humans and animals interact and pathogens are exchanged exist within the context of societal and policy interfaces (as depicted in the schematic in Figure 1). These networks exist at different scales. At a local-scale, households form part of what can be considered urban communities; groups of similar physical interfaces that are characterized by a set of societal (e.g., demographic and socioeconomic) characteristics. These communities are linked by the movement of people, livestock and their products, and wildlife, and the environment (which can conveniently be defined as networks of connectivity) [15,16]. As a result, key drivers that could promote interaction between humans and animals are: (i) livestock-keeping practices, production systems, and the movements of livestock and animal products in urban areas; and (ii) the direct effects of urbanization on the physical environment, ecosystems in which urban centers are developed, and animal communities that exist freely within these [8,17]. Urban systems are highly complex and the factors listed above are likely to influence the type and extent of human interactions with livestock, animal products, and ecosystems, resulting in the creating of human–animal interfaces that might promote the transmission of disease between animals and people.

Urban-adapted (referred to here as **synanthropic**) wildlife is abundant in cities, and is composed of species that can respond to behavioral and resource-based selection pressures

Glossary

Basic reproductive number (R0): the expected number of secondary cases produced by a single infection in a completely susceptible population. In order for a pathogen to spread and be maintained within a population of animals, the value of R0 must be >1.

Bridge host: an epidemiologically functional host population within the disease reservoir framework, which is able to transmit a pathogen from the maintenance community to a target population. To occupy this role, a host must satisfy the following two prerequisites: (i) either be competent for infection, replication and excretion of the pathogen but unable to maintain it alone, or be capable of mechanical transmission; and (ii) occupy an ecological niche that facilitates direct or indirect transmission between maintenance and target hosts.

Ecotone: edges or transitionary zones between adjacent ecological systems where biophysical factors, biological activity and ecological evolutionary processes are concentrated and intensified.

Emerging infectious disease (EID): either a newly recognized, clinically distinct infectious disease, or a known infectious disease whose reported incidence is increasing in a given place or within a specific population.

Interface (disease): a boundary across which parasites can be passed between biological communities. For our use of this term, an interface is defined by the community of species on both sides of the boundary (i.e., human–livestock–wildlife), and the biotic niches within which these communities exist.

Land-use change: changes to the structure of ecosystems as a result of human activities, which lead to perturbation of biotic systems. Examples include: deforestation, expansion of agriculture, pollution, depletion of marine fisheries, and eutrophication.

Maintenance host/community: the populations making up a disease reservoir. Maintenance hosts are species within which a pathogen can persist without reintroduction from another host, while a maintenance community is composed of epidemiologically linked populations

imposed by urban environments [18]. Many synanthropic species have been shown to carry zoonotic pathogens and in some cases act as reservoir hosts for these pathogens. Studies generally focus on those species that are found ubiquitously within human environments and that commonly act as hosts for zoonotic diseases, such as rodents, birds, bats, and certain other species of mammal (e.g., foxes in Europe and raccoons in the US) [19,20]. Rodents, for example, harbor important zoonoses such as plague, leptospirosis, and hantavirus infection, and the emergence and re-emergence of these pathogens in human populations is seemingly linked to increasing urbanization and urban poverty in developing countries and the ecology of zoonotic pathogens in rat populations [17,21–23]. Anthropogenic changes associated with urbanization can also bring bats into closer contact with livestock and humans and alter disease ecology [24,25]. As such, human activities that increase exposure to populations of urban-dwelling wildlife species will undoubtedly increase the risk of pathogens spilling over to humans or livestock, but little is known of the epidemiological processes by which this occurs at such interfaces.

Epidemiology at the Wildlife–Livestock–Human Interface

Most infectious agents circulate in communities composed of hosts that are infected with multiple parasites and parasites that can infect a variable diversity of hosts. Small changes in parasite community structure (within-host competition, or perturbations from host population dynamics) can result in far-reaching consequences for epidemiology of multihost and single host (monoxenous) parasite species [26–28]. Such downstream epidemiological effects are demonstrated in several well-studied zoonotic disease systems, including the seasonal and co-infection dynamics of cowpox virus [29], Lyme disease in white-footed mice [30], and Nipah and Hendra virus in fruit bats [24,31,32]. With the emergence of high-profile pathogens that exhibit wide host plasticity (such as Ebola and avian influenza viruses), a community approach is being increasingly embraced for studying the multihost ecology of zoonotic pathogens.

Studying the role of wildlife in multihost disease systems is complicated by ecological and behavioral attributes unique to these species, and the influence of natural and human systems; both of which complicate conceptual models of disease transmission [33]. Following the disease reservoirs framework recently revised by Viana *et al.* [34] and Caron *et al.* [35], in a multihost pathogen system where wildlife either exists within the **maintenance community** as a **maintenance host** or non-maintenance host, or outside the maintenance community as a **bridge host**, the dynamics of a zoonotic agent involve two phases: (i) transmission between maintenance and/or non-maintenance host species (wildlife and/or domestic) within the reservoir; and (ii) **spillover** transmission to humans from the maintenance community (Figure 1). In basic models, the persistence required for hosts to maintain a zoonotic pathogen and thus act as a maintenance community is determined by the **basic reproductive number** ($\mathbf{R_0}$: the transmission potential of a pathogen) and critical community size, while risk of spillover transmission to humans is defined by the force of infection from animals to humans. Contact is a key feature of both reservoir and disease emergence dynamics; R_0 is closely linked to the rate of contact between susceptible and infectious individuals and the recovery or mortality rate of infected individuals, and the force of infection (and thus risk of human spillover) is determined by prevalence of infection in the maintenance population and/or bridge hosts, the rate of contact between humans and infected individuals, and the probability that infection occurs upon contact [36–38]. However, host ecological traits (such as life-history characteristics, seasonality, coloniality, and sympatry) and population-level changes brought on by land-use change are likely to play a large role in pathogen transmission and persistence in wildlife and livestock species [33,39]. These factors (particularly human ecology) will strongly influence contact between wildlife, livestock, and humans, and prevalence of infection in animal reservoirs, and are therefore of fundamental importance to reservoir dynamics and disease emergence in changing landscapes.

within which a pathogen can persist indefinitely.
Network theory: the theory underlying network models. At their simplest, these are an adjacency matrix consisting of nodes (vertices) that represent individuals within a population, and edges (links) that represent interactions between individuals. In an epidemiological context, this provides a framework for visualising potential pathways of transmission within populations.
Population genetics (epidemiology): the study of the distribution and change in frequency of alleles within or between populations, and how the influences of selection, genetic drift, mutation, and gene flow are scaled to an individual, group, population, and landscape level. In doing so, researchers can assess the consequences of microevolutionary processes at differing scales.
Phylogenetics: the study of evolutionary relationships between individuals or species. These relationships are represented as a phylogeny (or evolutionary tree), consisting of a set of nodes (branching points) and edges.
Reservoir of infection: one or more epidemiologically connected populations in which a pathogen can be permanently maintained and from which infection is transmitted to a target population (such as humans).
Spillover: the disease dynamics that enable a pathogen to be transmitted into a susceptible target host population from its reservoir population.
Synanthropic wildlife: wildlife species that are ecologically associated with humans.
Target host: an epidemiologically functional host population within the disease reservoir framework, which is the focus for disease control.

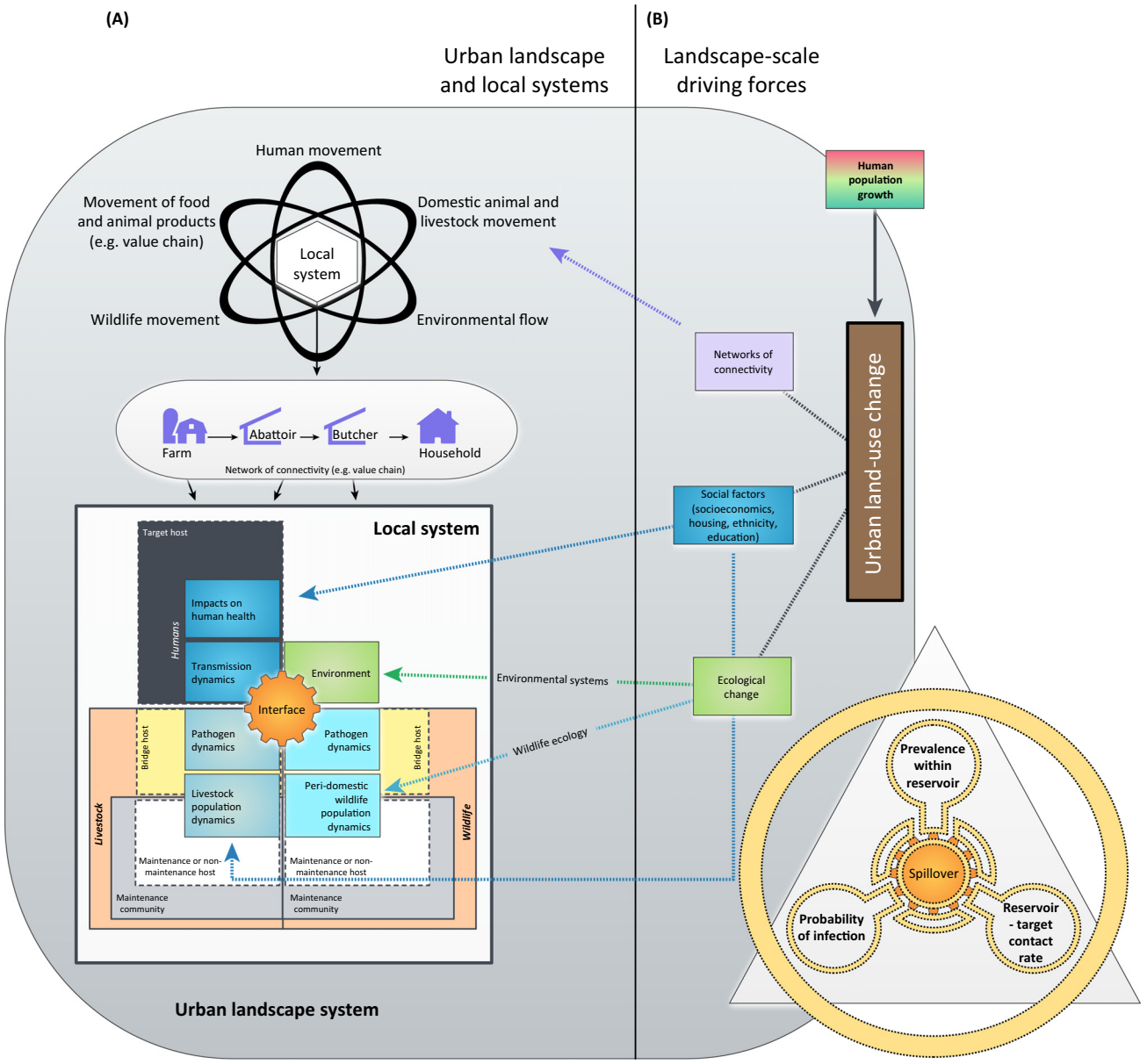


Figure 1. Conceptual Framework for Disease Emergence in Urban Landscapes (adapted, with permission, from [34,38]). It should be noted that we consider the structure of this framework as applicable to the emergence of antimicrobial resistance, as it is to disease emergence [92]. (A) This framework incorporates urban land-use change and its effects on two spatial scales: at a systems and local level. A simplified disease reservoir framework is included at the local level, in which livestock and synanthropic wildlife exist within the maintenance community as maintenance hosts (populations within the reservoir that can maintain the pathogen) or non-maintenance hosts (populations within the reservoir that cannot maintain the pathogen, therefore acting as vectors), or as bridge hosts that exist outside the maintenance community. (B) Following [38], spillover, which in this framework can relate to pathogen transfer in all directions except for target to reservoir, is governed by the force of infection consisting of the three elements shown.

Murray and Daszak [40] discuss two conceptual models for disease emergence under land-use change; the perturbation and pathogen pool hypotheses. The perturbation hypothesis focuses on a more dynamic model for disease emergence, where land-use change forces perturbations in pathogen dynamics within the reservoir, before emergence occurs in humans or livestock. The pathogen pool hypothesis assumes exposure to novel diseases from a diverse pool of

pathogens in wildlife to which humans or livestock, as naïve hosts, have not had prior exposure. In reality, it seems unlikely that these two hypotheses are mutually exclusive; evidence from empirical studies generally favors a dynamic model for disease emergence [41]. As such, the extent to which perturbation (changes in species richness, abundance, and contact rate) or the zoonotic pathogen pool dictate risk of emergence at urban interfaces, is probably dependent on the impact of urbanization on community ecology, and the degree of coevolution between sympatric wildlife, humans, and livestock at each interface.

Influence of Urbanization on Pathogen Dynamics within Multihost Wildlife Systems

Associations between urbanization and the prevalence of pathogens in populations of free-ranging wildlife have been described for a wide taxonomic range of host species and pathogens (reviewed in [13]). Evidence suggests that through altered habitat structure and changes to resource availability, urbanization results in significant changes to the structure of wildlife communities, which are subsequently characterized by low biodiversity with proportional increases in abundance of certain generalist species [42,43]. From a landscape-scale perspective, this results in a declining trend in species richness from rural areas to urban centers (biotic homogenization) with synanthropic species occurring at higher densities in urban and suburban environments than less-disturbed areas [13,44]. Not surprisingly, such profound changes in trophic structure will have epidemiological consequences for pathogens within these communities, and as a general rule, declining host biodiversity should be matched by a loss in parasite diversity, thus reducing the pathogen pool and with it the risk of novel disease emergence [45]. However, the epidemiological consequences of changes to such a system are likely to be pathogen specific, and dependent on how trophic reassortment affects the following parameters: likelihood of encounter and transmission between competent hosts, host abundance and/or density, and infected host mortality and recovery [46]. For example, helminth species richness of rodents in South East Asia is positively associated with decreasing rodent species richness, and increasing rodent abundance and level of synanthropy [47]. Increases in synanthropic species population density can elevate contact rates (through changes in host ranging patterns and densities), and thus increase the risk of pathogen transmission via direct contact and orofecal routes [37,44]. Fragmentation of these populations, in contrast, can result in genetic bottlenecks and subsequently reduced effective immune responses [48]. As host diversity decreases along gradients of urbanization, many pathogens are lost, but some (notably those in the hosts that remain in low diversity communities) can increase as a result of increased host abundance (termed the dilution effect) [30,49]. Reverse zoonotic transmission (zooanthroponosis) from humans to wildlife can also pose a threat to wildlife populations with increased exposure to humans [50,51]. The epidemiological effects of urbanization can therefore have important implications for both wildlife conservation and public health, with marginal wildlife species being susceptible to infection with pathogens circulating in urban-adapted hosts, and the potential for increased circulation of certain zoonotic pathogens in competent synanthropic reservoir hosts.

Interfaces between Sympatric Wildlife, Livestock and Humans in an Urban Landscape

Wildlife populations in urban landscapes are heterogeneously distributed, and certain species group in spatial aggregations with livestock (or their products) and humans, creating interfaces that might be important for the transmission of zoonotic agents. As described, the dynamics of infection at these interfaces are determined by changes in diversity, abundance and contact rates between reservoir and **target hosts**, thus influencing risk of cross-species pathogen transmission. Several systematic reviews have identified high-risk interfaces for zoonotic disease transmission on a global scale; specific interfaces for spillover from wildlife include human dwellings, agricultural fields, and occupational exposure, while broader descriptions include

Box 1. Dynamics of Cross-Species Transmission at an Urban Interface

In considering how urbanization could drive the emergence of a directly transmitted zoonotic agent through changes to the distributional ecology of wildlife, it is helpful to consider urban land-use change as a combination of abiotic factors that can influence the biotic niche within which a focal host species and pathogen exist. In a classical ecological sense, abiotic niches are used to describe a set of conditions that permit persistence of a host within a certain geographical range (such as climate), while the host species biotic niche is characterized by the inhibitory or facilitatory impact of other species on its existence [54]. Biotic factors are likely to be scale dependent; ecological and epidemiological processes can operate differently at different scales. For example, the prevalence of rodent-associated zoonoses varies widely between and within cities, likely being determined by site-specific abiotic factors such as physical microenvironments [17]. Efforts to understand scale-dependent processes, such as the role of geographical distribution and ecological setting in creating opportunities for pathogen transmission to occur, have led to the emergence of landscape epidemiology – a discipline that incorporates the framework of traditional epidemiology with the field of landscape ecology to facilitate the study of pathogens in relation to the ecosystem in which it is found and the human population at risk [55]. This field would have application when addressing the complex and profound effects of urbanization on wildlife population dynamics, and thus the ecology of zoonotic agents in these systems.

At a hypothetical wildlife–livestock–human interface one can imagine a cascade of abiotic and biotic changes creating conditions suitable for disease emergence at different scales. Abiotic factors (which include climate, resource provisioning, pollution, and habitat alteration) exist at multiple scales and facilitate or inhibit the survival of new and existing wildlife species within the reservoir community, which dictates the structural assemblage and fitness of hosts (see Figure 2 in main text). At a finer scale, these factors influence the biotic niche and dynamics of pathogens within the system; abiotic changes can directly influence microbiota (e.g., driving antimicrobial resistance) [56,57], whilst host diversity, density, phylogenetic structure (ecological, physiological and genetic similarity), and immunocompetence and immunological history of individuals all play an important role in host–pathogen interactions [58,59]. For example, not all conspecifics will be competent hosts for a given parasite, and as 'dead-end' hosts can play a role in regulating infection, while direct ecological interactions such as predation or competition will affect the population dynamics and distribution of competent reservoir hosts. Poulin [60] and Reperant [61] considered these factors as applied to the theory of island biogeography, where abiotic drivers influence the degree of interactions within source areas (sources of parasites such as wildlife reservoirs) and island areas (the recipient or target hosts), and the source-island distance (interactions between sources of parasites and recipient host populations that can drive spillover).

How these factors pertain to spillover to a target host (such as humans or livestock) is dependent on the force of infection; components of which include characteristics of the target host and the pathogen, and how the pathogen responds to changes in its biotic niche. Because abiotic factors influencing this system are driven by human activity, increased spatial overlap of humans, their livestock and wildlife is likely, but might not be enough to secure cross-species transmission. If the target host is human, then the likelihood of spillover can be moderated by individual and community variables such as social structure, living conditions, economic status, and health and risky behavior. From a pathogen genetic perspective, a jump into a new host species can either be nonadaptive (a process known as ecological fitting, where pre-existing traits allow emergence), or might require adaptive change (mutation of the pathogen in the reservoir host or the new host) [62]. Whether infecting new hosts is an adaptive or nonadaptive process depends on characteristics of the pathogen and host reservoir. Certain pathogens (RNA viruses in particular) seem inherently more adept at taking advantage of new epidemiological opportunities than others; possibly due to high mutation rates and broader host plasticity [52,63]. For others, structural properties of the reservoir can play a greater role and facilitate ecological fitting; phylogenetic distance between target (human) and reservoir hosts is a predictor of successful host jumps, while species belonging to certain phylogenetic clades might possess cellular components that make them more susceptible to pathogen invasion, regardless of phylogenetic distance from existing host species [52,64]. Benmayor *et al.* [65] showed that a higher density of susceptible hosts leads to an increased mutation rate and thus increased likelihood of viral host jumps occurring, while higher levels of interspecies transmission can lead to the adoption of more generalist pathogen virulence strategies [66]. Although conducted in unnatural microcosms, these studies demonstrate that the characteristics of the biotic niche of the pathogen, such as competition from sympatric microorganisms and host-reservoir dynamics, can also drive adaptive evolutionary processes. Finally, changes to population size and immune status of the reservoir can drive an increase in pathogen prevalence, thus amplifying the likelihood of spillover occurring. This leaves a key challenge to classify the levels of dynamic changes in organizational structure at different interfaces.

agricultural intensification and environmental change [8,52]. However, as argued by Jones *et al.* [8], attempts to describe systems within which pathogens emerge or change in virulence have predominantly focused on global generalizations, which might not be appropriate to capture the heterogeneity of interfaces. Instead, interfaces and the driving factors that define them should be studied at appropriate, spatially explicit scales [53]. We consider these feedback loops at hypothetical urban wildlife–livestock–human interfaces in Box 1.

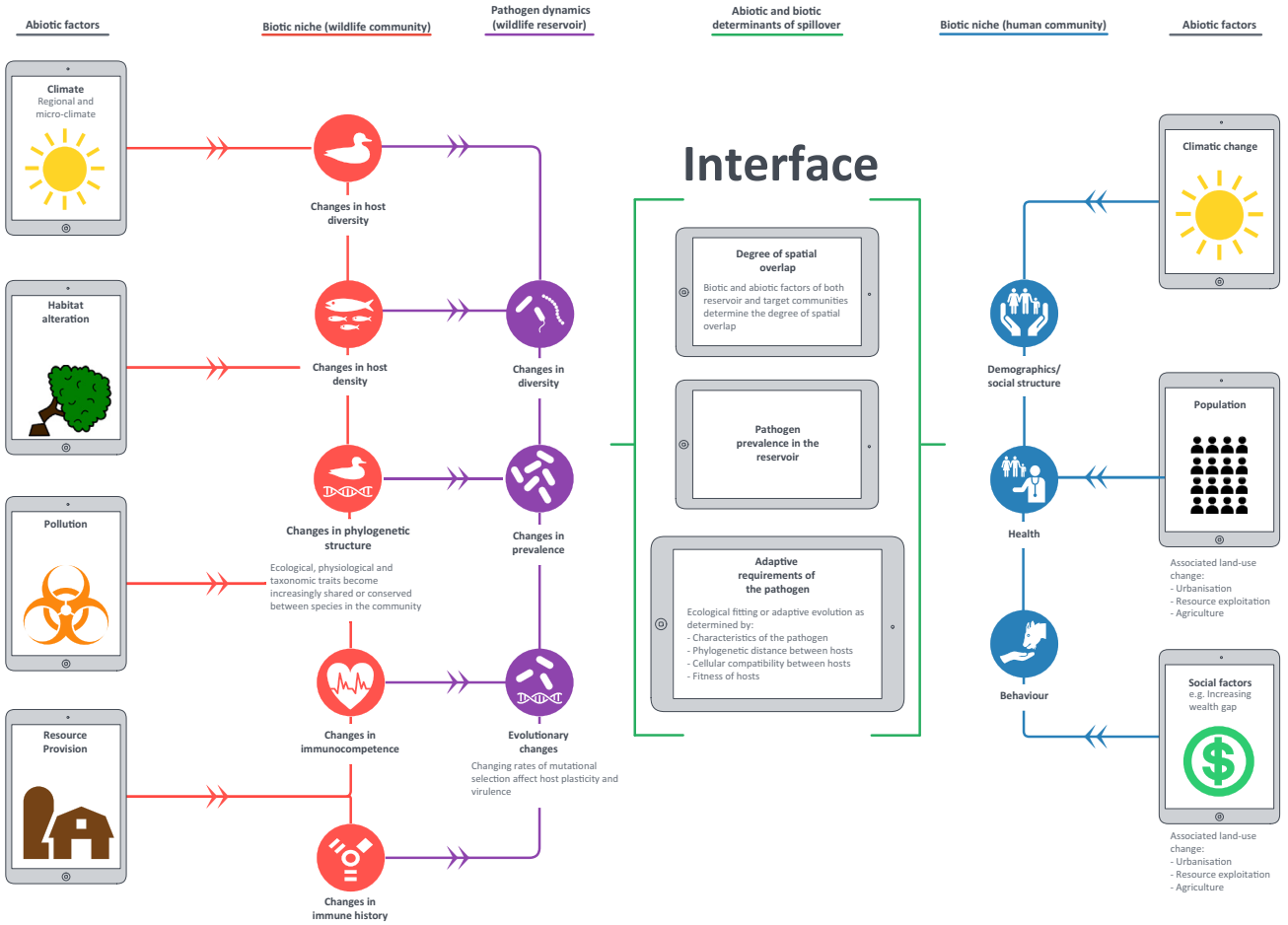


Figure 2. Cascades of Abiotic Factors, and the Components of Host–Pathogen Biotic System (for a Directly Transmitted Pathogen) That Are Affected by These Factors, Are Represented on Either Side of a Hypothetical Wildlife–Human Interface. Biotic systems represented here include a multihost wildlife community (acting as a parasite reservoir and composed of maintenance and non-maintenance hosts, and nonsusceptible species with direct ecological interactions with the reservoir), a human community (acting as variably susceptible target hosts) and the community of parasites within the wildlife reservoir. The requirements for spillover are represented centrally at the interface.

From an ecosystem perspective, anthropogenic pressures result in the fragmentation of natural biomes, leaving a composite mix of different habitats. Remnant fragments that are representative of the original biome can be thought of as patches that exist within a matrix of habitats that are unlike the original [67–69]. Interfaces between patches and the matrix exist at local-scales, and can be classified as **ecotones** – edges or transitionary zones between adjacent ecological systems where 'biophysical factors, biological activity and ecological evolutionary processes are concentrated and intensified' [70]. It has been suggested that by expanding ecotonal areas through interspersing human landscapes such as farmland and settlements with natural landscapes, anthropogenic influences can alter pathogen niches by bringing together humans, vectors, and reservoir hosts (wildlife or domestic animals), thus increasing contact and the risk of transmission [68]. Such landscape changes can be compounded by alterations in wildlife species interaction and abundance (e.g., host ecological traits); rodents can undergo ecological release at forest interfaces being attracted to farmland and human settlements for resources and suitable breeding habitat, and human settlements might provide suitable breeding habitat for mosquitos and birds (important arthropod vectors and reservoirs for West Nile virus) [70,71].

Evidence for an association between disease emergence and ecotones has been documented for several zoonoses with wildlife reservoirs, including yellow fever, Nipah virus encephalitis, influenza, rabies, hantavirus pulmonary syndrome, Lyme disease, cholera, *Escherichia coli* infection and African trypanosomiasis [70,72–74]. In urbanized areas such as cities, tangential variation in land use from rural–periurban–urban areas would be expected to generate a wide variety of ecotones on micro- and macrospatial scales. Ecotones can therefore represent important local-scale ecological interfaces within which zoonotic agents circulate and infect wildlife, domestic animals, and humans.

Another important factor in influencing interspecific wildlife contact and human–livestock–wildlife contact in urban environments is resource provisioning [19,20]. Clumping of resources occurs widely across urban environments at local (e.g., household) and landscape scales, whether as a result of variation in sanitation, refuse and agricultural byproducts, livestock-keeping practices, supplemental feeding of garden birds, or household food availability [13,75–78]. Informal livestock keeping is commonplace in African cities, and often characterized by low biosecurity and mixed-species livestock being kept in close proximity to humans. Evidence from recent zoonotic emergence events in Asia (such as Nipah and highly pathogenic avian influenza viruses) and the circulation of relatively stable zoonoses (such as hepatitis E and bovine tuberculosis) implicate a role for livestock acting as bridge hosts, epidemiologically linking wildlife and humans [31,79,80]. While resource provisioning commonly leads to increased contact rates between synanthropic wildlife, humans, and livestock, pathogen dynamics are also driven by susceptibility to infection, which, depending on the nature of provisioning, can be increased or decreased by host physical condition and immune defense [25,78,81]. In Eastern Australia, the decline in natural food resources and abundance of flowering resources in urban gardens has resulted in increasingly large urban colonies of *Pteropus* spp. bats (flying foxes) existing sympatrically with human and horse populations. These bats act as a reservoir for Hendra virus, and have historically lived in widely dispersed, interconnected metapopulations. Plowright *et al.* [25] demonstrated that the effects of urban development on these metapopulations, through increased contact with humans and horses, and reduced connectivity between flying fox colonies, could dramatically influence the epidemic dynamics of the virus in flying foxes, and increase the risk of Hendra virus emergence in horses and people. Using mechanistic models, Becker and Hall [82] and Becker *et al.* [78] also demonstrated host demographic, contact and immunological effects of provisioning on R_0, finding that unless provisioning reduces dietary exposure to pathogens or strongly improves host condition and immunity, increased aggregations of wildlife species dramatically increase pathogen invasion success and long-term prevalence. Environmental stressors such as heavy metal and pesticide pollutants, characteristic of certain urban environments, can further compound these outcomes through their effects on immunological function [83]. As such, resource provisioning is likely to increase host density (a key driver of transmission rates) and wildlife–livestock–human contact, making such areas important interfaces for disease emergence.

Table 1 applies the conceptual framework of wildlife–livestock–human interfaces developed by Jones *et al.* [8] to an urban setting such as Nairobi. Nairobi is a good example of a developing country city with human–livestock–synanthropic wildlife interfaces, and is a city, like many others, that has a growing boundary or edge which makes such contact more likely both on its edges and internally. In this context, we consider urban interfaces created through habitat fragmentation and resource provision. Such clear definition of interfaces is required to simplify the heterogeneous juxtaposition of humans and animals in urban landscapes, and thus enable the application of ecological, epidemiological, and anthropological approaches to the study of these landscapes. As well as capturing complex human and ecological processes that underlie disease emergence in urban landscapes, we believe that by studying these interfaces along rural–periurban–urban gradients, the landscape-level processes that accompany urbanization and underlie current theories of disease emergence could be captured.

Table 1. A Framework for Wildlife–Human–Livestock Interfaces in a Developing City such as Nairobi[a]

Description	Examples	Proposed level of wildlife–livestock–human contact
Urban ecotonal interfaces and fragmentation of natural ecosystems (anthropogenically derived habitat edges)	Forest edge; agricultural edge; incursions for natural resource harvesting; urban wetlands	Increasing contact between humans, livestock and wildlife (both nonsynanthropic and synanthropic species)
Evolving urban landscape – areas of informally planned resource provision	Informal refuse dumps; increasingly intensive farming and associated value chains (low biosecurity); backyard farming	High contact between humans, livestock and synanthropic wildlife that is largely unmanaged
Managed urban landscape – areas of formally planned resource provision	Sewage plants; established intensive farming and associated value chains (high biosecurity)	Controlled contact between humans and livestock Little contact between wildlife, livestock, and humans
Managed urban landscape – areas of recreational habitat suitable for wildlife	Parks and recreation facilities; gardens	Few contacts between humans and livestock, and livestock and wildlife Increasing contact between humans and synanthropic wildlife

[a]Adapted from a broader conceptual framework describing types of wildlife–livestock–human interface and their characteristics, developed by Jones *et al.* [8].

Concluding Remarks and Future Directions

In this review, we consider the role that urbanization plays in the emergence of zoonoses, through exploring the ecological complexity of wildlife–livestock–human interfaces. In doing so we argue that interfaces should be considered a critical component of disease ecology in changing urban landscapes, and echo a body of recent literature calling for greater ecological sophistication in epidemiological theories of disease emergence [84–87]. The majority of epidemiological studies use foundational concepts to study a single, or small number of well-characterized host species and pathogens when investigating transmission and connectivity within multihost systems. While this approach is well established, and useful in developing frameworks upon which the empirical characterization of a known host–pathogen system can be determined (through mechanistic models) and interventions planned (e.g., [34]), focus on a single species or pathogen might hinder the detection of pathogen emergence within a structurally complex system by overshadowing the evolutionary and transmission processes that precede this. As signaled by the emerging field of community disease ecology (reviewed in [87]), new approaches are required to investigate disease emergence, that shift focus from the pathogen to understanding the processes underlying emergence [35]. In response, disease ecologists have moved towards adopting principles from community ecology; including metapopulation and **network theory**, trait-based approaches and a consideration of processes acting across biological scales [27,53,84,86–89]. The development of new modeling techniques will play a key role, and several frameworks have been suggested, that focus on integrating broad methodologies and crossdisciplinary collaborations to investigate causation in disease emergence [53,90,91]. Such methods will be key to unraveling the structural complexity of ecological communities at wildlife–livestock–human interfaces, and thus understanding how they function as epidemiological systems prior to disease emergence.

While the focus of this review is on disease emergence, we would like to highlight the relevance of the frameworks discussed in combination with the broader concept of urban interfaces, for studying antimicrobial resistance (AMR). Currently considered urgent One Health issues, it is likely that the emergence of AMR and zoonotic pathogens in urban areas are underlined by a similar set of societal and ecological drivers [92]. Given the current rate of urbanization, and potential for associated changes in societal structure, food systems, and natural ecosystems to

Outstanding Questions

We consider the following unresolved questions as central to shedding light on the complex set of conditions required for a pathogen to enter a new host. Such studies will contribute to the development of more realistic mechanistic frameworks for cross-species spillover, and the design of appropriate interventions and control strategies.

Characterization of interfaces

At which urban animal–animal and animal–human interfaces is spillover of priority zoonotic pathogens most likely to occur?

What are the forces driving the creation of these interfaces?

What role does the environment and environmental change play in the transmission and spillover risk for zoonotic pathogens at these interfaces? How does this vary across gradients of urbanization?

Interface dynamics

Reservoir communities and intermediate (bridge) hosts

Wildlife reservoirs represent complex communities of maintenance and non-maintenance hosts, and conspecifics that could have a regulatory effect on parasite dynamics through ecological interactions with hosts. How are wildlife species assembled at proposed high-risk urban wildlife–livestock–human interfaces, and how does this vary across gradients of urbanization?

What is the presence and prevalence of zoonoses in urban synanthropic wildlife, and how does this vary across gradients of urbanization?

How are multispecies wildlife communities epidemiologically structured at high-risk urban wildlife–livestock–human interfaces, and how does this vary across gradients of urbanization?

What is the role of wildlife in contributing to genetic pools of antibiotic resistance across urban landscapes?

How does urban land-use change affect host fitness and immunity in synanthropic wildlife and livestock

Box 2. Future Perspective: Linking Community and Disease Ecology through Molecular Epidemiology

We consider how molecular epidemiology and network theory could provide a platform from which to investigate epidemiological connectivity, by mapping transmission and detecting pathogen adaptation across multiple scales of biological organization, and in doing so add real-world complexity to the study of disease emergence at interfaces.

Studying Structural Complexity Using Networks

By considering each species that exists within the community of interest as part of a network, it is possible to assess how species attributes (such as parasite diversity, preferred habitat, or social rank) vary within and between communities. In particular, the application of network theory to study epidemiology within ecological networks promises to address some of the shortcomings of approaches such as mathematical modeling and experimental studies (i.e., reductionist, lacking in biological reality), that are traditionally used to understand how pathogens behave in host communities. By capturing the structural complexity and heterogeneous mixing of individuals within a population, epidemiological networks can be used to investigate factors affecting transmission, while also providing a realistic framework for modeling pathogen spread through the community [94–96]. However, the structure of epidemiological networks within multihost pathogen systems remains largely uninvestigated; the majority of studies that have applied network analysis to an epidemiological system for free-ranging wildlife consider transmission as a function of observed interactions between individuals belonging to a single species.

As initially demonstrated by VanderWaal *et al.* [97], molecular tools allow researchers to move beyond the assumption that observed contacts are reflective of transmission, by offering the potential to generate quantifiable measures of transmission between individuals of the same or different species. This has permitted studies to investigate community drivers of shared-parasite transmission; both VanderWaal and Atwill [98] and Blyton *et al.* [99] found covariance between networks of shared *Escherichia coli* genotypes and social contacts in giraffes and possums, respectively, such that individuals that were centrally located within social networks also acted as hubs of transmission. Gene dispersal measures represent just one **population genetics** approach to apportioning genetic variation that occurs within and between different populations, and have been successful to inform epidemiological connectivity between humans, wildlife, and livestock [72,77]. Others have used networks to describe population-level drivers in parasite diversity; Anthony *et al.* [100] developed network models of virus families in rhesus macaques, demonstrating that viral community assembly exhibits nonrandom patterns, which suggests that the effect of deterministic factors on viral diversity should be predictable. Thus, providing the limitations of a network approach are fully considered (reviewed in [94]), these tools could be used to describe the structure of wildlife communities implicated in zoonotic transmission to livestock and/or humans, identify key drivers that influence risk of transmission (such as land-use change or climate), and thus assist in untangling the complexity of epidemiological processes at interfaces in a realistic manner, regardless of the taxonomic distance between hosts.

Phylogenetics

Phylogenetic approaches focus on the study of evolutionary relationships among genetic lineages and can be used to reconstruct epidemiological histories from pathogen genetic sequence data. Given appropriate metadata (e.g., the date and host from which the sequence was sampled) it has been possible to reconstruct most likely populations of origin, historical host-switching events, and transmission pathways over long periods of time [101]. One potential approach for studying how ecology and spatial distribution of hosts affects the transmission and evolution of their parasites is to test for the association of host or environmental traits (such as species ecological characteristics, geographic location, behavior, or physical characteristics) with the phylogenetic structure of pathogen genetic sequences obtained from these hosts. For example, Parker *et al.* [102] developed a Bayesian Markov-Chain Monte Carlo approach for testing whether closely related taxa are more likely to share a trait of interest. This tool has proved useful in describing the spatial distribution of a number of pathogens (e.g., distribution patterns of yellow fever virus in Venezuela and evolutionary spread of influenza viruses in migratory birds [103,104]), and similar techniques have been embraced in community ecology [105]. Although the application of ecological trait-mapping onto phylogenies to understand how abiotic and biotic factors relate to parasite phylogenetic structure has not been explored for disease ecology, its potential is recognized by Suzán *et al.* [89].

By utilizing both population genetic (gene dispersal measures and network models) and phylogenetic trait-mapping approaches to investigate the epidemiological structure of multihost wildlife communities, it should be possible to depict epidemiological connectivity at an individual, intracommunity, and intercommunity scale. These techniques are therefore very appealing for determining connectivity within reservoir populations and between reservoir and target hosts at wildlife–livestock–human interfaces, and might provide an opportunity to inform the most appropriate targets for surveillance and control.

expose human and animal populations to novel pathogens, we recommend an interdisciplinary approach to studying urban human–wildlife–livestock interfaces, with the following aims: (i) establish characterizations for potential high-risk interfaces that exist along gradients of urbanization, and identify processes that have led to their formation; (ii) describe biological organization

populations, and what effect (if any) does this have on circulating zoonotic pathogens?

How does microbial diversity in wildlife species (i.e., the pathogen pool) vary according to urban land-use change?

Determinants of spillover

Can pathogen sequence data shed light on adaptive and nonadaptive evolutionary processes occurring as pathogens are transmitted between species at urban interfaces? How do pathogen evolutionary processes relate to phylogenetic distance between reservoir, bridge, and target host species?

How does direct and indirect contact between wildlife, livestock and humans vary under differing livestock management conditions, and in response to broader biotic and abiotic factors in urban environments (e.g., anthropogenic behaviour, socioeconomic status, species diversity and climatic variation)?

What are the finer-scale epidemiological connections between synanthropic wildlife, livestock, humans, and their shared environments, and how is the risk of zoonotic pathogen transmission influenced by human and wildlife traits in urban environments (e.g., anthropogenic behavior, socioeconomic status, species diversity, and climatic variation)?

How does urban land-use change affect host fitness and immunity in synanthropic wildlife, livestock and human populations, and what effect (if any) does this have on circulating zoonotic pathogens?

and community ecology at these interfaces, conduct surveillance for priority zoonotic pathogens (i.e., those with emergent potential) across host taxa, and study the evolutionary processes underlying cross-species transmission where it is detected (see Box 2); and (iii) at interfaces where transmission risks are identified, develop appropriate interventions that can be used to reduce risk of transmission. Given their epidemiological significance, interfaces represent a critical point of control for the transmission of zoonoses. A detailed discussion of control measures is beyond the scope of this article, but interventions could be implemented at an interface (i.e., preventative action such as husbandry and behavioral changes) or policy level (for a complete review, see [93]). If, as we discuss in this review, pathogen dynamics at interfaces are characterized by dynamic changes in community structure driven by abiotic factors, emphasis should be focused on studying epidemiological connectivity (i.e., pathways and heterogeneity of transmission – see Box 2) and how this changes longitudinally with time. Such studies will be crucial in identifying the dynamic processes responsible for driving changes in community structure and thus pathogen dynamics at different interfaces over time (see Outstanding Questions).

Acknowledgments

We acknowledge the valuable insight and critical feedback provided by three anonymous reviewers during the review process. This article was supported by the UK Medical Research Council, the Natural Environment Research Council (UK), the Economic and Social Research Council (UK), and the Biotechnology and Biosciences Research Council (UK), through the Environmental & Social Ecology of Human Infectious Diseases Initiative (ESEI), Grant Reference: G1100783/1. This work also received support from the CGIAR Research Program on Agriculture for Nutrition and Health (A4NH), led by the International Food Policy Research Institute (IFPRI). We also acknowledge the CGIAR Fund Donors (http://www.cgiar.org/who-we-are/cgiar-fund/fund-donors-2). M.J.W. was supported by a Sir Henry Wellcome Postdoctoral Fellowship from the Wellcome Trust (WT103953MA).

References

1. Smolinsky, M.S. *et al.* (2003) . In *Microbial Threats to Health: Emergence, Detection, and Response* (Smolinski, M.S. *et al.*, eds), National Academies Press
2. Woolhouse, M.E.J. and Gowtage-Sequeria, S. (2005) Host range and emerging and reemerging pathogens. *Emerg. Infect. Dis.* 11, 1842–1847
3. Jones, K.E. *et al.* (2008) Global trends in emerging infectious diseases. *Nature* 451, 990–993
4. Karesh, W.B. *et al.* (2012) Ecology of zoonoses: natural and unnatural histories. *Lancet* 380, 1936–1945
5. McMichael, J. (2004) Environmental and social influences on emerging infectious diseases: past, present and future. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 359, 1049–1058
6. Daszak, P. (2000) Emerging Infectious diseases of wildlife-threats to biodiversity and human health. *Science* 287, 443–449
7. Perry, B.D. *et al.* (2011) Livestock and global change special feature: current drivers and future directions of global livestock disease dynamics. *Proc. Natl. Acad. Sci. U.S.A.* 110, 20871–20877
8. Jones, B.A. *et al.* (2013) Zoonosis emergence linked to agricultural intensification and environmental change. *Proc. Natl. Acad. Sci. U.S.A.* 110, 8399–8404
9. Patz, J.A. *et al.* (2004) Unhealthy landscapes: policy recommendations on land use change and infectious disease emergence. *Environ. Health Perspect.* 112, 1092–1098
10. McFarlane, R.A. *et al.* (2013) Land-use change and emerging infectious disease on an island continent. *Int. J. Environ. Res. Public Health* 10, 2699–2719
11. Knobler, S. *et al.* (2006) The impact of globalization on infectious disease emergence and control: exploring the consequences and opportunities. In *Proceedings of the Forum on Microbial Threats* (Knobler, S. *et al.*, eds), pp. 21–48, Institute of Medicine
12. Nations, U. (2014) *World Urbanization Prospects: The 2014 Revision,* Department of Economic and Social Affairs, Population Division, (New York)
13. Bradley, C.A. and Altizer, S. (2007) Urbanization and the ecology of wildlife diseases. *Trends Ecol. Evol.* 22, 95–102
14. Alirol, E. *et al.* (2011) Urbanisation and infectious diseases in a globalised world. *Lancet Infect. Dis.* 11, 131–141
15. Alberti, M. (2005) The effects of urban patterns on ecosystem function. *Int. Reg. Sci. Rev.* 28, 168–192
16. Pickett, S.T.A. *et al.* (2011) Urban ecological systems: scientific foundations and a decade of progress. *J. Environ. Manage.* 92, 331–362
17. Himsworth, C.G. *et al.* (2013) Rats, cities, people, and pathogens: a systematic review and narrative synthesis of literature regarding the ecology of rat-associated zoonoses in urban centers. *Vector Borne Zoonotic Dis.* 13, 349–359
18. Lowry, H. *et al.* (2012) Behavioural responses of wildlife to urban environments. *Biol. Rev. Camb. Philos. Soc.* 88, 537–549
19. Deplazes, P. *et al.* (2004) Wilderness in the city: The urbanization of *Echinococcus multilocularis*. *Trends Parasitol.* 20, 77–84
20. Wright, A.N. and Gompper, M.E. (2005) Altered parasite assemblages in raccoons in response to manipulated resource availability. *Oecologia* 144, 148–156
21. Guan, P. *et al.* (2009) Investigating the effects of climatic variables and reservoir on the incidence of hemorrhagic fever with renal syndrome in Huludao City, China: a 17-year data analysis based on structure equation model. *BMC Infect. Dis.* 9, 109
22. Evangelista, K. and Coburn, J. (2010) *Leptospira* as an emerging pathogen: a review of its biology, pathogenesis, and host immune responses. *Future Microbiol.* 5, 1413–1425
23. Halliday, J.E.B. *et al.* (2013) Urban leptospirosis in Africa: a cross-sectional survey of *Leptospira* infection in rodents in the Kibera urban settlement, Nairobi, Kenya. *Am. J. Trop. Med. Hyg.* 89, 1095–1102
24. Pulliam, J.R.C. *et al.* (2012) Agricultural intensification, priming for persistence and the emergence of Nipah virus: a lethal bat-borne zoonosis. *J. R. Soc. Interface* 9, 89–101
25. Plowright, R.K. *et al.* (2011) Urban habituation, ecological connectivity and epidemic dampening: the emergence of Hendra virus from flying foxes (*Pteropus* spp.). *Proc. R. Soc. B* 278, 3703–3712

26. Hall, S.R. *et al.* (2005) Selective predation and productivity jointly drive complex behavior in host–parasite systems. *Am. Nat.* 165, 70–81
27. Telfer, S. *et al.* (2010) Species interactions in a parasite community drive infection risk in a wildlife population. *Science* 330, 243–246
28. Tomczuk, K. *et al.* (2015) Analysis of intrinsic and extrinsic factors influencing the dynamics of bovine *Eimeria* spp. from central–eastern Poland. *Vet. Parasitol.* 214, 22–28
29. Begon, M. *et al.* (2009) Seasonal host dynamics drive the timing of recurrent epidemics in a wildlife population. *Proc. R. Soc. B* 276, 1603–1610
30. Keesing, F. *et al.* (2006) Effects of species diversity on disease risk. *Ecol. Lett.* 9, 485–498
31. Daszak, P. *et al.* (2006) The emergence of Nipah and Hendra virus: pathogen dynamics across a wildlife-livestock-human continuum. In *Disease Ecology: Community Structure and Pathogen Dynamics* (Collinge, S. and Ray, C., eds), pp. 186–201, Oxford University Press
32. Plowright, R.K. *et al.* (2015) Ecological dynamics of emerging bat virus spillover. *Proc. R. Soc. B* 282 http://dx.doi.org/10.1098/rspb.2014.2124
33. Hayman, D.T.S. *et al.* (2013) Ecology of zoonotic infectious diseases in bats: current knowledge and future directions. *Zoonoses Public Health* 60, 2–21
34. Viana, M. *et al.* (2014) Assembling evidence for identifying reservoirs of infection. *Trends Ecol. Evol.* 29, 270–279
35. Caron, A. *et al.* (2015) Bridge hosts, a missing link for disease ecology in multi-host systems. *Vet. Res.* 46, 83
36. Begon, M. *et al.* (2002) A clarification of transmission terms in host-microparasite models: numbers, densities and areas. *Epidemiol. Infect.* 129, 147–153
37. Davis, S. and Calvet, E. (2005) Fluctuating rodent populations and risk to humans from rodent-borne zoonoses. *Vector-Borne Zoonotic Dis.* 5, 305–314
38. Lloyd-Smith, J.O. *et al.* (2009) Epidemic dynamics at the human–animal interface. *Science* 326, 1362–1367
39. Gardner-Santana, L.C. *et al.* (2009) Commensal ecology, urban landscapes, and their influence on the genetic characteristics of city-dwelling Norway rats (*Rattus norvegicus*). *Mol. Ecol.* 18, 2766–2778
40. Murray, K. and Daszak, P. (2013) Human ecology in pathogenic landscapes: two hypotheses on how land use change drives viral emergence. *Curr. Opin. Virol.* 3, 79–83
41. Lloyd-Smith, J.O. *et al.* (2014) Nine challenges in modelling the emergence of novel pathogens. *Epidemics* 10, 35–39
42. Faeth, S.H. *et al.* (2005) Trophic dynamics in urban communities. *Bioscience* 55, 399
43. Galbraith, J.A. *et al.* (2015) Supplementary feeding restructures urban bird communities. *Proc. Natl. Acad. Sci. U.S.A.* 112, e2648–e2657
44. Evans, K.L. *et al.* (2009) Habitat influences on urban avian assemblages. *Ibis* 151, 19–39
45. Hudson, P.J. *et al.* (2006) Is a healthy ecosystem one that is rich in parasites? *Trends Ecol. Evol.* 21, 381–385
46. Wood, C.L. *et al.* (2014) Does biodiversity protect humans against infectious disease? *Ecology* 95, 817–832
47. Chaisiri, K. *et al.* (2010) Human-dominated habitats and helminth parasitism in Southeast Asian murids. *Parasitol. Res.* 107, 931–937
48. Sommer, S. (2005) The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Front. Zool.* 2, 16–33
49. Ezenwa, V.O. *et al.* (2006) Avian diversity and West Nile virus: testing associations between biodiversity and infectious disease risk. *Proc. R. Soc. B* 273, 109–117
50. Palacios, G. *et al.* (2011) Human metapneumovirus infection in wild mountain gorillas, Rwanda. *Emerg. Infect. Dis.* 17, 711–713
51. Franklinos, L.H.V. *et al.* (2015) *Streptococcus pyogenes* infection in a free-living European hedgehog (*Erinaceus europaeus*). *Ecohealth* 12, 689–692
52. Kreuder Johnson, C. *et al.* (2015) Spillover and pandemic properties of zoonotic viruses with high host plasticity. *Sci. Rep.* 5, 14830
53. Wood, J.L.N. *et al.* (2012) A framework for the study of zoonotic disease emergence and its drivers: spillover of bat pathogens as a case study. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 367, 2881–2892
54. Estrada-Peña, A. *et al.* (2014) Effects of environmental change on zoonotic disease risk: an ecological primer. *Trends Parasitol.* 30, 205–214
55. Meentemeyer, R.K. *et al.* (2012) Landscape epidemiology of emerging infectious diseases in natural and human-altered ecosystems. *Annu. Rev. Phytopathol.* 50, 379–402
56. Olson, D.H. *et al.* (2013) Mapping the global emergence of *Batrachochytrium dendrobatidis*, the amphibian chytrid fungus. *PLoS ONE* 8 http://dx.doi.org/10.1371/journal.pone.0056802
57. Wellington, E.M.H. *et al.* (2013) The role of the natural environment in the emergence of antibiotic resistance in gram-negative bacteria. *Lancet Infect. Dis.* 13, 155–165
58. Dobson, A. (2004) Population dynamics of pathogens with multiple host species. *Am. Nat.* 164, 64-S78
59. Pilosof, S. *et al.* (2014) Host-parasite network structure is associated with community-level immunogenetic diversity. *Nat. Commun.* 5, 5172
60. Poulin, R. (2004) Macroecological patterns of species richness in parasite assemblages. *Basic Appl. Ecol.* 5, 423–434
61. Reperant, L.A. (2010) Applying the theory of island biogeography to emerging pathogens: toward predicting the sources of future emerging zoonotic and vector-borne diseases. *Vector Borne Zoonotic Dis.* 10, 105–110
62. Pepin, K.M. *et al.* (2010) Identifying genetic markers of adaptation for surveillance of viral host jumps. *Nat. Rev. Microbiol.* 8, 802–813
63. Woolhouse, M.E.J. *et al.* (2005) Emerging pathogens: the epidemiology and evolution of species jumps. *Trends Ecol. Evol.* 20, 238–244
64. Longdon, B. *et al.* (2011) Host phylogeny determines viral persistence and replication in novel hosts. *PLoS Pathog.* http://dx.doi.org/10.1371/journal.ppat.1002260
65. Benmayor, R. *et al.* (2009) Host mixing and disease emergence. *Curr. Biol.* 19, 764–767
66. Gandon, S. (2004) Evolution of multihost parasites. *Evolution* 58, 455–469
67. Ellis, E.C. and Ramankutty, N. (2008) Putting people in the map: anthropogenic biomes of the world. *Front. Ecol. Environ.* 6, 439–447
68. Reisen, W.K. (2010) Landscape epidemiology of vector-borne diseases. *Annu. Rev. Entomol.* 55, 461–483
69. Wilcove, D.S. *et al.* (1986) Habitat fragmentation in the temperate zone. In *Conservation Biology: The Science of Scarcity and Diversity* (Soulé, E., ed.), pp. 237–256, Sinaure Associates
70. Despommier, D. *et al.* (2006) The role of ecotones in emerging infectious diseases. *Ecohealth* 3, 281–289
71. Brown, H.E. *et al.* (2008) Remotely-sensed vegetation indices identify mosquito clusters of West Nile virus vectors in an urban landscape in the northeastern United States. *Vector Borne Zoonotic Dis.* 8, 197–206
72. Goldberg, T. *et al.* (2008) Forest fragmentation as cause of bacterial transmission among nonhuman primates, humans, and livestock, Uganda. *Emerg Infect. Dis.* 14, 1375–1382
73. Eisen, R.J. *et al.* (2007) A spatial model of shared risk for plague and hantavirus pulmonary syndrome in the southwestern United States. *Am. J. Trop. Med. Hyg.* 77, 999–1004
74. Hahn, M.B. *et al.* (2014) The role of landscape composition and configuration on *Pteropus giganteus* roosting ecology and Nipah virus spillover risk in Bangladesh. *Am. J. Trop. Med. Hyg.* 90, 247–255
75. Jobbins, S.E. *et al.* (2014) Leptospira interrogans at the human–wildlife interface in northern Botswana: a newly identified public health threat. *Zoonoses Public Health* 61, 113–123
76. Lawson, B. *et al.* (2014) Epidemiological evidence that garden birds are a source of human salmonellosis in England and Wales. *PLoS ONE* http://dx.doi.org/10.1371/journal.pone.0088968

77. Swirski, A.L. *et al.* (2014) Spatial epidemiology of *Escherichia coli* O157:H7 in dairy cattle in relation to night roosts of Sturnus vulgaris (European starling) in Ohio, USA (2007-2009). *Zoonoses Public Health* 61, 427–435

78. Becker, D.J. *et al.* (2015) Linking anthropogenic resources to wildlife-pathogen dynamics: a review and meta-analysis. *Ecol. Lett.* 18, 483–495

79. Sims, L.D. *et al.* (2005) Origin and evolution of highly pathogenic H5N1 avian influenza in Asia. *Vet. Rec.* 157, 159–164

80. Meng, X.J. *et al.* (2009) Wild boars as sources for infectious diseases in livestock and humans. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 364, 2697–2707

81. Murray, M. *et al.* (2015) Poor health is associated with use of anthropogenic resources in an urban carnivore. *Proc. R. Soc. B* 282 http://dx.doi.org/10.1098/rspb.2015.0009

82. Becker, D. and Hall, R. (2014) Too much of a good thing: resource provisioning alters infectious disease dynamics in wildlife. *Biol. Lett.* 10 http://dx.doi.org/10.1098/rsbl.2014.0309

83. Padgett, D.A. and Glaser, R. (2003) How stress influences the immune response. *Trends Immunol.* 24, 444–448

84. Roche, B. *et al.* (2012) Linking community and disease ecology: the impact of biodiversity on pathogen transmission. *Philos. Trans. R. Soc. B Biol. Sci.* 367, 2807–2813

85. Dunne, J.A. *et al.* (2013) Parasites affect food web structure primarily through increased diversity and complexity. *PLoS Biol.* http://dx.doi.org/10.1371/journal.pbio.1001579

86. Rushmore, J. *et al.* (2013) Social network analysis of wild chimpanzees provides insights for predicting infectious disease risk. *J. Anim. Ecol.* 82, 976–986

87. Johnson, P.T.J. *et al.* (2015) Why infectious disease research needs community ecology. *Science* 349 http://dx.doi.org/10.1126/science.1259504

88. Luis, A.D. *et al.* (2013) A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special? *Proc. Biol. Sci. B* 280 http://dx.doi.org/10.1098/rspb.2012.2753

89. Suzán, G. *et al.* (2015) Metacommunity and phylogenetic structure determine wildlife and zoonotic infectious disease patterns in time and space. *Ecol. Evol.* 5, 865–873

90. Restif, O. *et al.* (2012) Model-guided fieldwork: practical guidelines for multidisciplinary research on wildlife ecological and epidemiological dynamics. *Ecol. Lett.* 15, 1083–1094

91. Buhnerkempe, M.G. *et al.* (2015) Eight challenges in modelling disease ecology in multi-host, multi-agent systems. *Epidemics* 10, 26–30

92. Robinson, T.P. *et al.* (2016) Antibiotic resistance is the quintessential One Health issue. *Trans. R. Soc. Trop. Med. Hyg.* 7, 377–380

93. Gortazar, C. *et al.* (2015) The wild side of disease control at the wildlife–livestock–human interface: a review. *Front. Vet. Sci.* 1, 1–12

94. Godfrey, S.S. (2013) Networks and the ecology of parasite transmission: a framework for wildlife parasitology. *Int. J. Parasitol. Parasites Wildl.* 2, 235–245

95. Leventhal, G.E. *et al.* (2015) Evolution and emergence of infectious diseases in theoretical and real-world networks. *Nat. Commun.* 2, 1–11

96. Pilosof, S. *et al.* (2015) Potential parasite transmission in multihost networks based on parasite sharing. *PLoS ONE* http://dx.doi.org/10.1371/journal.pone.0117909

97. VanderWaal, K.L. *et al.* (2014) Quantifying microbe transmission networks for wild and domestic ungulates in Kenya. *Biol. Conserv.* 169, 136–146

98. VanderWaal, K. and Atwill, E. (2014) Linking social and pathogen transmission networks using microbial genetics in giraffe (*Giraffa camelopardalis*). *J. Anim. Ecol.* 83, 406–414

99. Blyton, M.D.J. *et al.* (2014) Not all types of host contacts are equal when it comes to *E. coli* transmission. *Ecol. Lett.* 17, 970–978

100. Anthony, S.J. *et al.* (2015) Non-random patterns in viral diversity. *Nat. Commun.* 6, 1–7

101. Archie, E.A. *et al.* (2009) Infecting epidemiology with genetics: a new frontier in disease ecology. *Trends Ecol. Evol.* 24, 21–30

102. Parker, J. *et al.* (2008) Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty. *Infect. Genet. Evol.* 8, 239–246

103. Scotch, M. *et al.* (2014) Diffusion of influenza viruses among migratory birds with a focus on the Southwest United States. *Infect. Genet. Evol.* 26, 185–193

104. Auguste, A.J. *et al.* (2015) Enzootic transmission of yellow fever virus, Venezuela. *Emerg. Infect. Dis.* 21, 99–102

105. Cavender-Bares, J. *et al.* (2009) The merging of community ecology and phylogenetic biology. *Ecol. Lett.* 12, 693–715